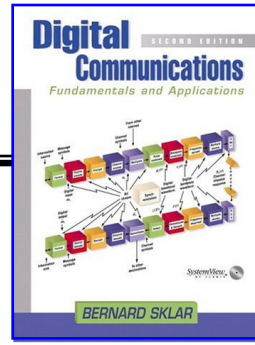


DIGITAL COMMUNICATIONS

Fundamentals and Applications

Second Edition

BERNARD SKLAR

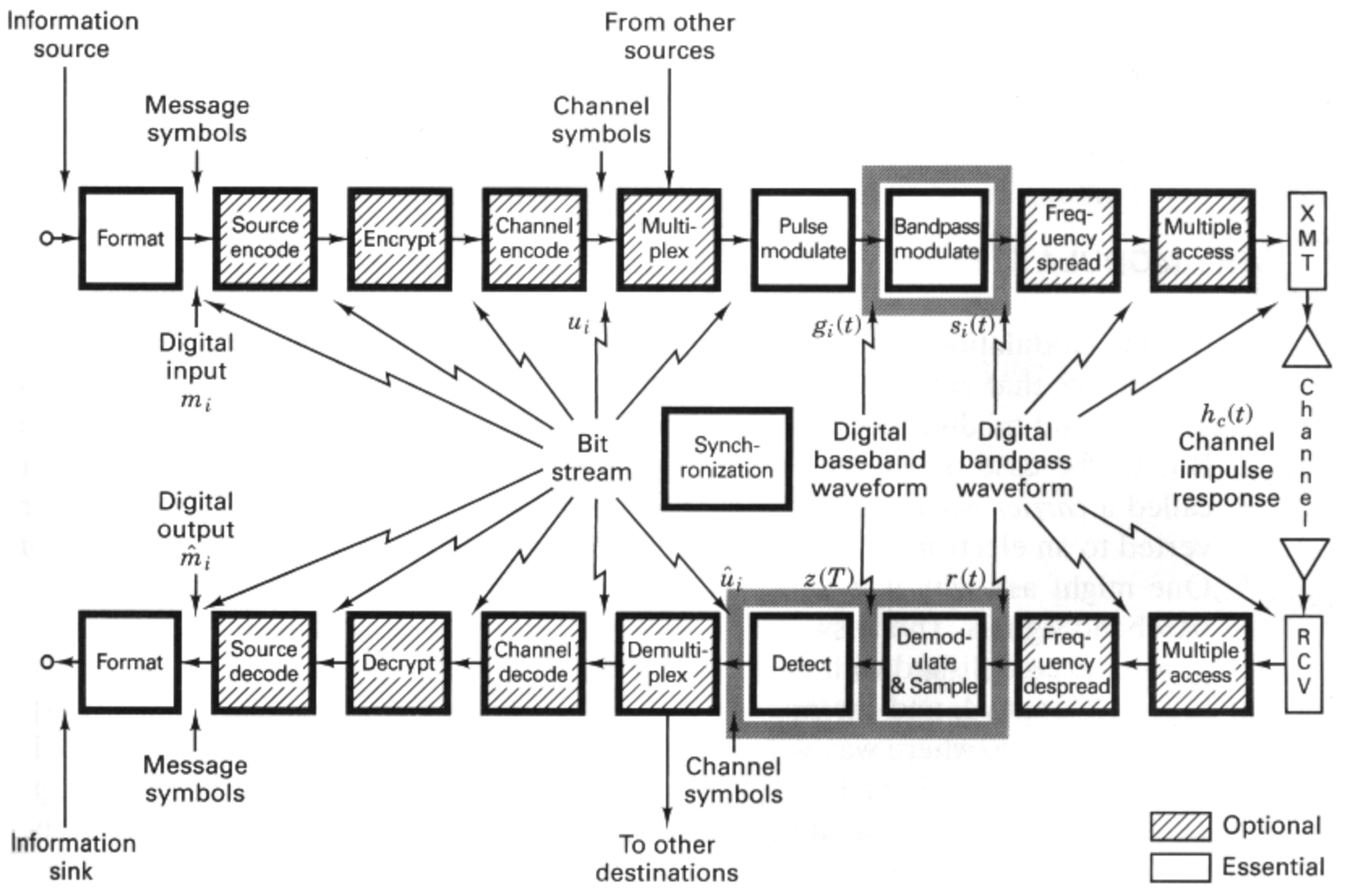
Communications Engineering Services, Tarzana, California
and
University of California, Los Angeles

PH PTR

Prentice Hall P T R
Upper Saddle River, New Jersey 07458
www.phptr.com

ISBN 0-13-084788-7
90000
9 780130 847881

Bandpass Modulation and Demodulation



4.1 WHY MODULATE?

Digital modulation is the process by which digital symbols are transformed into waveforms that are compatible with the characteristics of the channel. In the case of baseband modulation, these waveforms usually take the form of shaped pulses. But in the case of *bandpass modulation* the shaped pulses modulate a sinusoid called a *carrier wave,* or simply a *carrier;* for radio transmission the carrier is converted to an electromagnetic (EM) field for propagation to the desired destination. One might ask why it is necessary to use a carrier for the radio transmission of baseband signals. The answer is as follows. The transmission of EM fields through space is accomplished with the use of antennas. The size of the antenna depends on the wavelength λ and the application. For cellular telephones, antennas are typically $\lambda/4$ in size, where wavelength is equal to c/f, and c, the speed of light, is 3×10^8 m/s. Consider sending a baseband signal (say, $f = 3000$ Hz) by coupling it to an antenna directly without a carrier wave. How large would the antenna have to be? Let us size it by using the telephone industry benchmark of $\lambda/4$ as the antenna dimension. For the 3,000 Hz baseband signal, $\lambda/4 = 2.5 \times 10^4$ $m \approx 15$ miles. To transmit a 3,000 Hz signal through space, *without carrier-wave modulation,* an antenna that spans 15 miles would be required. However, if the baseband information is first modulated on a higher frequency carrier, for example a 900 MHz carrier, the equivalent antenna diameter would be about 8 cm. For this reason, carrier-wave or bandpass modulation is an essential step for all systems involving radio transmission.

Bandpass modulation can provide other important benefits in signal transmission. If more than one signal utilizes a single channel, modulation may be used to separate the different signals. Such a technique, known as *frequency-division multiplexing,* is discussed in Chapter 11. Modulation can be used to minimize the effects of interference. A class of such modulation schemes, known as *spread-spectrum modulation,* requires a system bandwidth much larger than the minimum bandwidth that would be required by the message. The trade-off of bandwidth for interference rejection is considered in Chapter 12. Modulation can also be used to place a signal in a frequency band where design requirements, such as filtering and amplification, can be easily met. This is the case when radio-frequency (RF) signals are converted to an intermediate frequency (IF) in a receiver.

4.2 DIGITAL BANDPASS MODULATION TECHNIQUES

Bandpass modulation (either analog or digital) is the process by which an information signal is converted to a sinusoidal waveform; for digital modulation, such a sinusoid of duration T is referred to as a digital symbol. The sinusoid has just three features that can be used to distinguish it from other sinusoids: amplitude, frequency, and phase. Thus bandpass modulation can be defined as the process whereby the amplitude, frequency, or phase of an RF carrier, or a combination of them, is varied in accordance with the information to be transmitted. The general form of the carrier wave is

$$s(t) = A(t) \cos \theta(t) \tag{4.1}$$

where $A(t)$ is the time-varying amplitude and $\theta(t)$ is the time-varying angle. It is convenient to write

$$\theta(t) = \omega_0 t + \phi(t) \tag{4.2}$$

so that

$$s(t) = A(t) \cos [\omega_0 t + \phi(t)] \tag{4.3}$$

where ω_0 is the *radian frequency* of the carrier and $\phi(t)$ is the *phase.* The terms f and ω will each be used to denote frequency. When f is used, frequency in hertz is intended; when ω is used, frequency in radians per second is intended. The two frequency parameters are related by $\omega = 2\pi f$.

The basic *bandpass modulation/demodulation* types are listed in Figure 4.1. When the receiver exploits knowledge of the carrier's phase to detect the signals, the process is called *coherent detection;* when the receiver does not utilize such phase reference information, the process is called *noncoherent detection.* In digital communications, the terms *demodulation* and *detection* are often used interchangeably, although demodulation emphasizes waveform recovery, and detection entails the process of symbol decision. In ideal coherent detection, there is available at the receiver a prototype of each possible arriving signal. These prototype waveforms attempt to duplicate the transmitted signal set in every respect, even RF phase. The

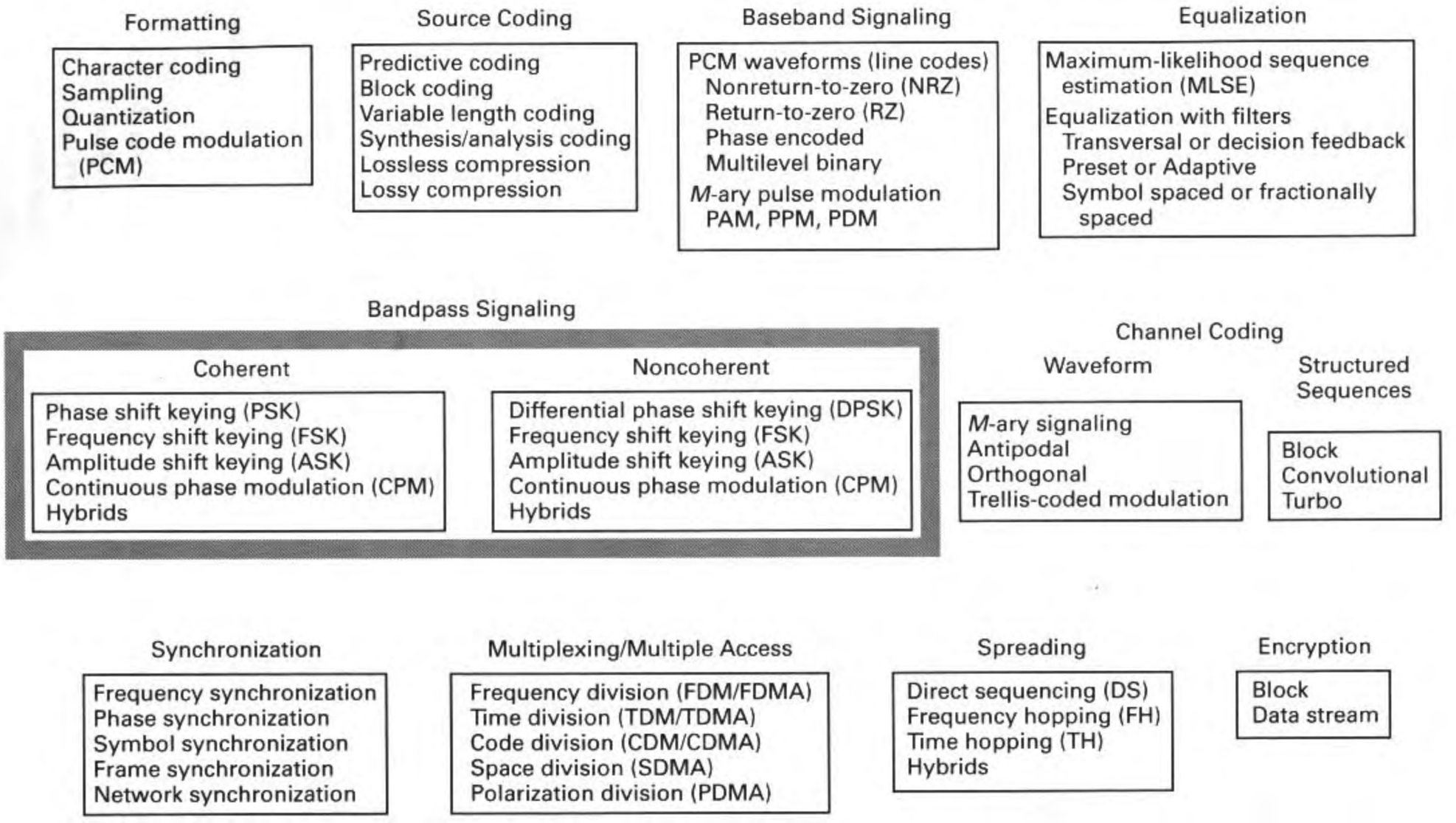


Figure 4.1 Basic digital communication transformations.

receiver is then said to be *phase locked* to the incoming signal. During demodulation, the receiver multiplies and integrates (correlates) the incoming signal with each of its prototype replicas. Under the heading of coherent modulation/demodulation in Figure 4.1 are listed phase shift keying (PSK), frequency shift keying (FSK), amplitude shift keying (ASK), continuous phase modulation (CPM), and hybrid combinations. The basic bandpass modulation formats are discussed in this chapter. Some specialized formats, such as offset quadrature PSK (OQPSK), minimum shift keying (MSK) belonging to the CPM class, and quadrature amplitude modulation (QAM), are treated in Chapter 9.

Noncoherent demodulation refers to systems employing demodulators that are designed to operate without knowledge of the absolute value of the incoming signal's phase; therefore, phase estimation is not required. Thus the advantage of noncoherent over coherent systems is reduced complexity, and the price paid is increased probability of error (P_E). In Figure 4.1 the modulation/demodulation types that are listed in the noncoherent column, DPSK, FSK, ASK, CPM, and hybrids, are similar to those listed in the coherent column. We had implied that phase information is not used for noncoherent reception; how do you account for the fact that there is a form of phase shift keying under the noncoherent heading? It turns out that an important form of PSK can be classified as noncoherent (or differentially coherent) since it does not require a reference in phase with the received carrier. This "pseudo-PSK," termed *differential PSK* (DPSK), utilizes phase information of the prior symbol as a phase reference for detecting the current symbol. This is described in Sections 4.5.1 and 4.5.2.

4.2.1 Phasor Representation of a Sinusoid

Using a well-known trigonometric identity called Euler's theorem, we introduce the complex notation of a sinusoidal carrier wave as follows:

$$e^{j\omega_0 t} = \cos \omega_0 t + j \sin \omega_0 t \tag{4.4}$$

One might be more comfortable with the simpler, more straightforward notation $\cos \omega_0 t$ or $\sin \omega_0 t$. What possible benefit can there be with the complex notation? We will see (in Section 4.6) that this notation facilitates our description of how real-world modulators and demodulators are implemented. For now, let us point to the general benefits of viewing a carrier wave in the complex form of Equation (4.4).

First, within this compact from, $e^{j\omega_0 t}$, is contained the two important quadrature components of any sinusoidal carrier wave, namely the inphase (real) and the quadrature (imaginary) components that are orthogonal to each other. Second, the unmodulated carrier wave is conveniently represented in a polar coordinate system as a unit vector or phasor rotating counterclockwise at the constant rate of ω_0 radians/s, as depicted in Figure 4.2. As time is increasing (i.e., from t_0 to t_1) we can visualize the time-varying projections of the rotating phasor on the inphase (I) axis and the quadrature (Q) axis. These cartesian axes are usually referred to as the I channel and Q channel respectively, and the projections on them represent the

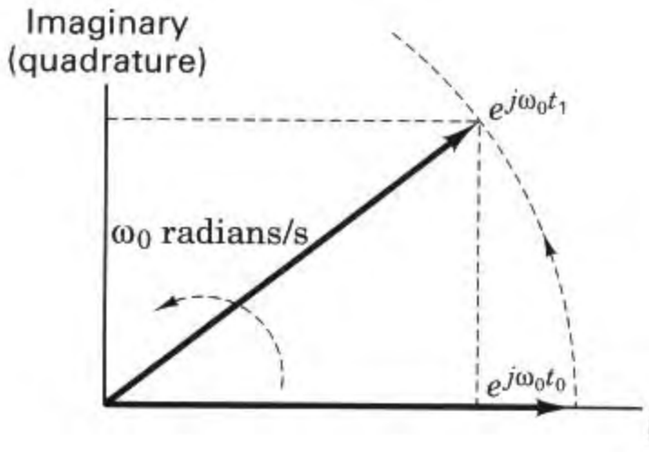


Figure 4.2 Phasor representation of a sinusoid.

signal components (orthogonal to each other) associated with those channels. Third, when it comes time to modulate the carrier wave with information, we can view this modulation as a methodical perturbation of the rotating phasor (and its projections).

For example, consider a carrier wave that is *amplitude modulated* (AM) with a sinusoid having an amplitude of unity and a frequency ω_m, where $\omega_m \ll \omega_0$. The analytical form of the transmitted waveform is

$$s(t) = \mathrm{Re}\left\{e^{j\omega_0 t}\left(1 + \frac{e^{j\omega_m t}}{2} + \frac{e^{-j\omega_m t}}{2}\right)\right\} \tag{4.5}$$

where $\mathrm{Re}\{x\}$ is the real part of the complex quantity $\{x\}$. Figure 4.3 illustrates that the rotating phasor $e^{j\omega_0 t}$ of Figure 4.2 is now perturbed by two sideband terms—$e^{j\omega_m t}/2$ rotating counterclockwise and $e^{-j\omega_m t}/2$ rotating clockwise. The sideband phasors are rotating at a much slower speed than the carrier-wave phasor. The net result of the composite signal is that the rotating carrier-wave phasor now appears to be growing longer and shorter pursuant to the dictates of the sidebands, but its frequency stays constant—hence, the term "amplitude modulation."

Another example to reinforce the usefulness of the phasor view is that of *frequency modulating* (FM) the carrier wave with a similar sinusoid having a frequency of ω_m radians/s. The analytical representation of *narrowband* FM (NFM) has an appearance similar to AM and is represented by

$$s(t) = \mathrm{Re}\left\{e^{j\omega_0 t}\left(1 - \frac{\beta}{2}e^{-j\omega_m t} + \frac{\beta}{2}e^{j\omega_m t}\right)\right\} \tag{4.6}$$

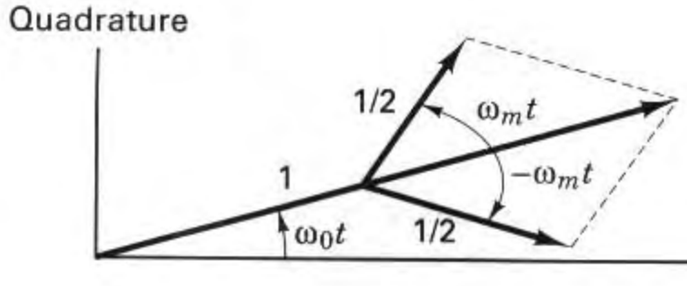


Figure 4.3 Amplitude modulation.

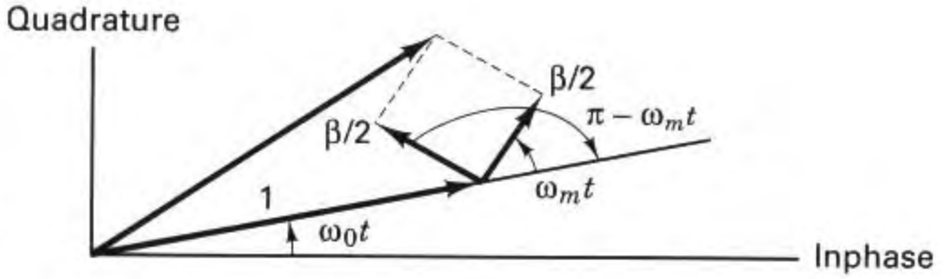


Figure 4.4 Narrowband frequency modulation.

where β is the modulation index [1]. Figure 4.4 illustrates that the rotating carrier-wave phasor is again perturbed by two sideband terms, but because one of the sideband terms carries a minus sign in Equation (4.6), the clockwise and counterclockwise rotating sideband phasors have a different symmetry than in the case of AM. In the case of AM the sideband symmetry results in the carrier-wave phasor growing longer and shorter with time. In NFM, the sideband symmetry (90° different than AM) results in the carrier-wave phasor speeding up and slowing down according to the dictates of the sidebands, but the amplitude says essentially constant—hence, the term "frequency modulation."

Figure 4.5 illustrates examples of the most common digital modulation formats: PSK, FSK, ASK, and a hybrid combination of ASK and PSK (ASK/PSK or APK). The first column lists the analytic expression, the second is a typical pictorial of the waveform versus time, and the third is a vector (or phasor) schematic, with the orthogonal axes labeled $\{\psi_j(t)\}$. In the general M-ary signaling case, the processor accepts k source bits (or channel bits if there is coding) at a time and instructs the modulator to produce one of an available set of $M = 2^k$ waveform types. Binary modulation, where $k = 1$, is just a special case of M-ary modulation.

In Figure 4.2, we represented a carrier wave as a phasor rotating in a plane at the speed of the carrier-wave frequency ω_0 radians/s. In Figure 4.5, the phasor schematic for each digital-modulation example represents a constellation of information signals (vectors or points in the signaling space), where time is not represented. In other words, the constantly rotating aspect of the unmodulated carrier wave has been removed, and only the information-bearing phasor positions, relative to one another, are presented. Each example in Figure 4.5 uses a particular value of M, the set size.

4.2.2 Phase Shift Keying

Phase shift keying (PSK) was developed during the early days of the deep-space program; PSK is now widely used in both military and commercial communications systems. The general analytic expression for PSK is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left[\omega_0 t + \phi_i(t)\right] \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.7}$$

where the phase term, $\phi_i(t)$, will have M discrete values, typically given by

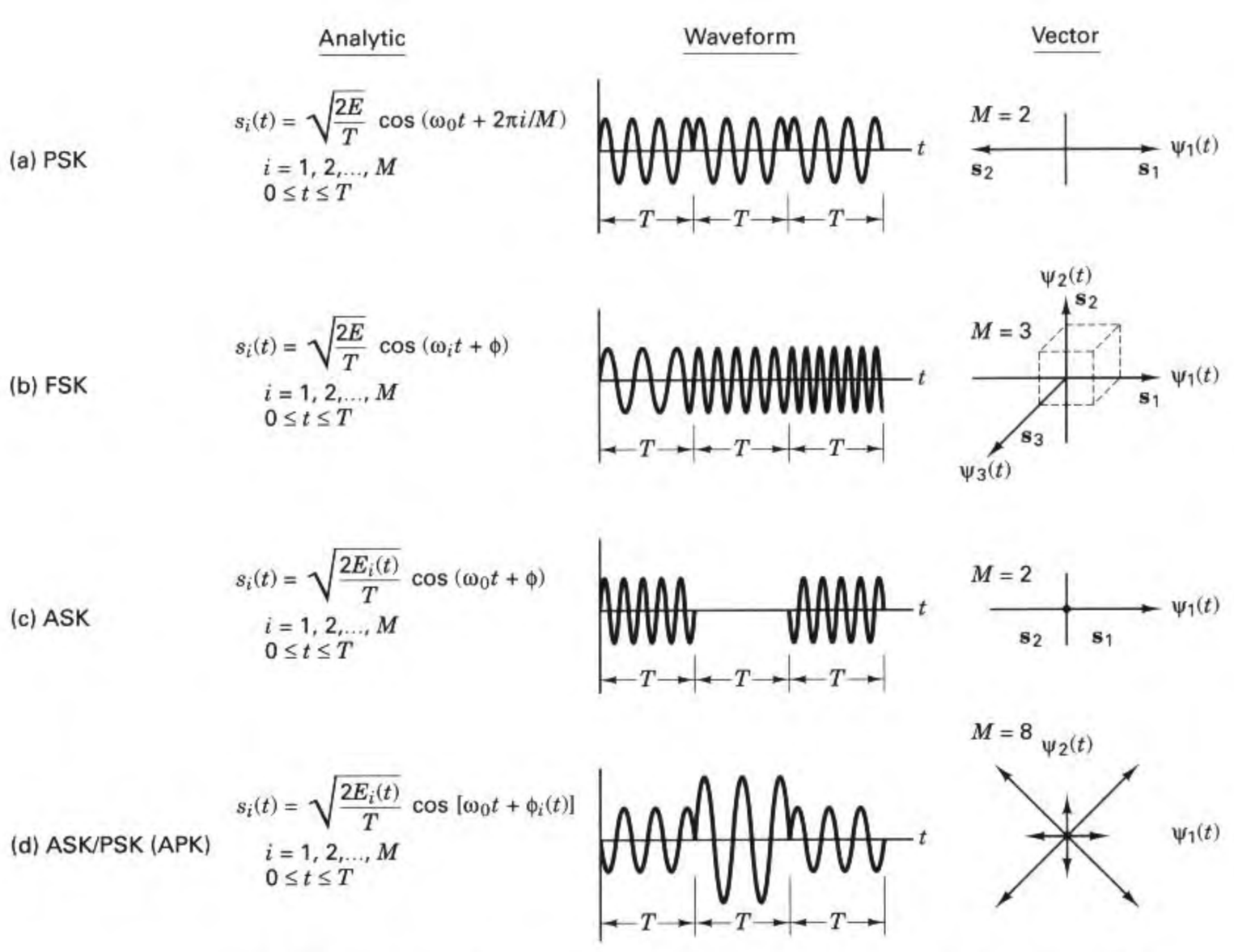


Figure 4.5 Digital modulations. (a) PSK. (b) FSK. (c) ASK. (d) ASK/PSK (APK).

$$\phi_i(t) = \frac{2\pi i}{M} \qquad i = 1, \ldots, M$$

For the binary PSK (BPSK) example in Figure 4.5a, M is 2. The parameter E is symbol energy, T is symbol time duration, and $0 \le t \le T$. In BPSK modulation, the modulating data signal shifts the phase of the waveform $s_i(t)$ to one of two states, either zero or π (180°). The waveform sketch in Figure 4.5a shows a typical BPSK waveform with its abrupt phase changes at the symbol transitions; if the modulating date stream were to consist of alternating ones and zeros, there would be such an abrupt change at each transition. The signal waveforms can be represented as vectors or phasors on a polar plot; the vector length corresponds to the signal amplitude, and the vector direction for the general M-ary case corresponds to the signal phase relative to the other $M - 1$ signals in the set. For the BPSK example, the vector picture illustrates the two 180° opposing vectors. Signal sets that can be depicted with such opposing vectors are called *antipodal signal sets.*

4.2.3 Frequency Shift Keying

The general analytic expression for FSK modulation is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.8}$$

where the frequency term ω_i has M discrete values, and the phase term ϕ is an arbitrary constant. The FSK waveform sketch in Figure 4.5b illustrates the typical frequency changes at the symbol transitions. At the symbol transitions, the figure depicts a gentle shift from one frequency (tone) to another. This behavior is only true for a special class of FSK called continuous-phase FSK (CPFSK) which is described in Section 9.8. In the general MFSK case, the change to a different tone can be quite abrupt, because there is no requirement for the phase to be continuous. In this example, M has been chosen equal to 3, corresponding to the same number of waveform types (3-ary); note that this $M = 3$ choice for FSK has been selected to emphasize the mutually perpendicular axes. In practice, M is usually a nonzero power of 2 (2, 4, 8, 16, ...). The signal set is characterized by Cartesian coordinates, such that each of the mutually perpendicular axes represents a sinusoid with a different frequency. As described earlier, signal sets that can be characterized with such mutually perpendicular vectors are called *orthogonal* signals. Not all FSK signaling is orthogonal. For any signal set to be orthogonal, it must meet the criterion set forth in Equation (3.69). For an FSK signal set, in the process of meeting this criterion, a condition arises on the spacing between the tones in the set. The necessary frequency spacing between tones to fulfill the orthogonality requirement is discussed in Section 4.5.4.

4.2.4 Amplitude Shift Keying

For the ASK example in Figure 4.5c, the general analytic expression is

$$s_i(t) = \sqrt{\frac{2E_i(t)}{T}} \cos(\omega_0 t + \phi) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.9}$$

where the amplitude term $\sqrt{2E_i(t)/T}$ will have M discrete values, and the phase term ϕ is an arbitrary constant. In Figure 4.5c, M has been chosen equal to 2, corresponding to two waveform types. The ASK waveform sketch in the figure can describe a radar transmission example, where the two signal amplitude states would be $\sqrt{2E/T}$ and zero. The vector picture utilizes the same phase–amplitude polar coordinates as the PSK example. Here we see a vector corresponding to the maximum-amplitude state, and a point at the origin corresponding to the zero-amplitude state. Binary ASK signaling (also called on–off keying) was one of the earliest forms of digital modulation used in radio telegraphy at the beginning of this century. Simple ASK is no longer widely used in digital communications systems, and thus it will not be treated in detail here.

4.2.5 Amplutude Phase Keying

For the combination of ASK and PSK (APK) example in Figure 4.5d, the general analytic expression

$$s_i(t) = \sqrt{\frac{2E_i(t)}{T}} \cos\left[\omega_0 t + \phi_i(t)\right] \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.10}$$

illustrates the indexing of both the signal amplitude term and the phase term. The APK waveform picture in Figure 4.5d illustrates some typical simultaneous phase and amplitude changes at the symbol transition times. For this example, M has been chosen equal to 8, corresponding to eight waveforms (8-ary). The figure illustrates a hypothetical eight-vector signal set on the phase-amplitude plane. Four of the vectors are at one amplitude, and the other four vectors are at a different amplitude. Each of the vectors is separated by 45°. When the set of M symbols in the two-dimensional signal space are arranged in a rectangular constellation, the signaling is referred to as quadrature amplitude modulation (QAM); examples of QAM are considered in Chapter 9.

The vector picture for each of the modulation types described in Figure 4.5 (except the FSK case) is characterized on a plane whose *polar* coordinates represent signal *amplitude* and *phase.* The FSK case assumes orthogonal FSK (see Section 4.5.4) and is characterized in a *Cartesian* coordinate space, with each axis representing a *frequency tone* ($\cos \omega_i t$) from the M-ary set of orthogonal tones.

4.2.6 Waveform Amplitude Coefficient

The waveform amplitude coefficient appearing in Equations (4.7) to (4.10) has the same general form $\sqrt{2E/T}$ for all modulation formats. The derivation of this expression begins with

$$s(t) = A \cos \omega t \tag{4.11}$$

where A is the peak value of the waveform. Since the peak value of a sinusoidal waveform equals $\sqrt{2}$ times the root-mean-square (rms) value, we can write

$$\begin{aligned} s(t) &= \sqrt{2} A_{\text{rms}} \cos \omega t \\ &= \sqrt{2A_{\text{rms}}^2} \cos \omega t \end{aligned}$$

Assuming the signal to be a voltage or a current waveform, A_{rms}^2 represents average power P (normalized to 1 Ω). Therefore, we can write

$$s(t) = \sqrt{2P} \cos \omega t \tag{4.12}$$

Replacing P watts by E joules/T seconds, we get

$$s(t) = \sqrt{\frac{2E}{T}} \cos \omega t \tag{4.13}$$

We shall use either the amplitude notation A in Equation (4.11) or the designation $\sqrt{2E/T}$ in Equation (4.13). Since the *energy* of a received signal is the key parameter in determining the error performance of the detection process, it is often more convenient to use the amplitude notation in Equation (4.13) because it facilitates solving directly for the probability of error P_E as a function of signal energy.

4.3 DETECTION OF SIGNALS IN GAUSSIAN NOISE

The bandpass model of the detection process is virtually identical to the baseband model considered in Chapter 3. That is because a received bandpass waveform is first transformed to a baseband waveform before the final detection step takes place. For linear systems, the mathematics of detection is unaffected by a shift in frequency. In fact, we can define an *equivalence theorem* as follows: Performing bandpass linear signal processing, followed by heterodyning the signal to baseband yields the same results as heterodyning the bandpass signal to baseband, followed by baseband linear signal processing. The term "heterodyning" refers to a frequency *conversion* or *mixing* process that yields a spectral shift in the signal. As a result of this equivalence theorem, all linear signal-processing simulations can take place at baseband (which is preferred for simplicity), with the same results as at bandpass. This means that the performance of most digital communication systems will often be described and analyzed as if the transmission channel is a baseband channel.

4.3.1 Decision Regions

Consider that the two-dimensional signal space in Figure 4.6 is the locus of the noise-perturbed prototype binary vectors $(\mathbf{s}_1 + \mathbf{n})$ and $(\mathbf{s}_2 + \mathbf{n})$. The noise vector, $\mathbf{n}$, is a zero-mean random vector; hence the received signal vector, $\mathbf{r}$, is a random vector with mean $\mathbf{s}_1$ or $\mathbf{s}_2$. The detector's task after receiving $\mathbf{r}$ is to decide which of the signals ($\mathbf{s}_1$ or $\mathbf{s}_2$,) was actually transmitted. The method is usually to decide on the signal classification that yields the minimum expected P_E, although other strategies are possible [2]. For the case where M equals 2, with $\mathbf{s}_1$ and $\mathbf{s}_2$ being equally likely and with the noise being an additive white Gaussian noise (AWGN) process, we will see that the minimum-error decision rule is equivalent to choosing the signal class such that the distance $d(\mathbf{r}, \mathbf{s}_i) = \| \mathbf{r} - \mathbf{s}_i \|$ is minimized, where $\| \mathbf{x} \|$ is called the *norm* or *magnitude* of vector $\mathbf{x}$. This rule is often stated in terms of decision regions. In Figure 4.6, let us construct decision regions in the following way. Draw a line connecting the tips of the prototype vectors $\mathbf{s}_1$ and $\mathbf{s}_2$. Next, construct the perpendicular bisector of the connecting line. Notice that this bisector passes through the origin of the space if $\mathbf{s}_1$ and $\mathbf{s}_2$ are equal in amplitude. For this $M = 2$ example, in Figure 4.6, the constructed perpendicular bisector represents the locus of points equidistant between $\mathbf{s}_1$ and $\mathbf{s}_2$; hence, the bisector describes the boundary between decision region 1 and decision region 2. The *decision rule* for the detector, stated in terms of *decision regions,* is as follows: Whenever the received signal $\mathbf{r}$ is located

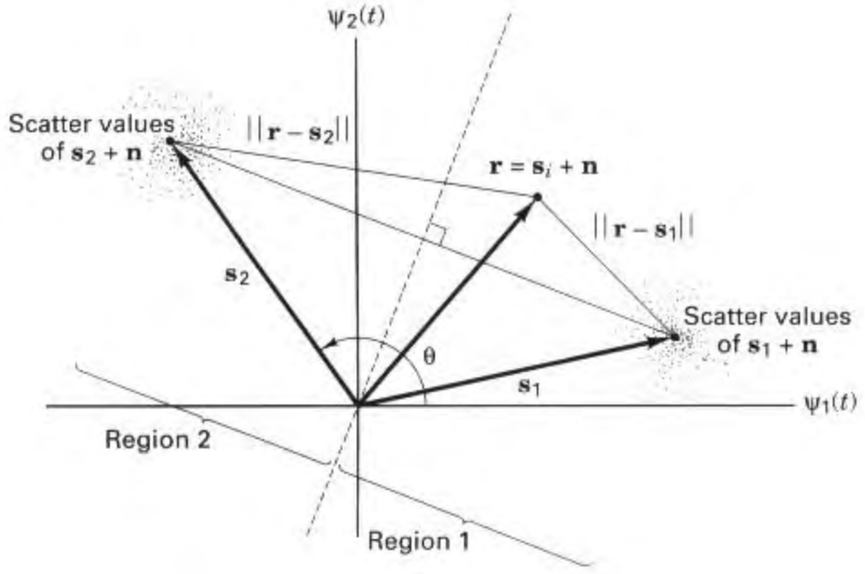


Figure 4.6 Two-dimensional signal space, with arbitrary equal-amplitude vectors $\mathbf{s}_1$ and $\mathbf{s}_2$.

in region 1, choose signal $\mathbf{s}_1$; when it is located in region 2, choose signal $\mathbf{s}_2$. In Figure 4.6, if the angle θ equals 180°, then the signal set $\mathbf{s}_1$ and $\mathbf{s}_2$ represents BPSK. However, in this figure, θ was purposely chosen to be less than 180° in order to emphasize the idea of decision regions in general.

4.3.2 Correlation Receiver

In Section 4.2 we treated the detection of *baseband* binary signals in Gaussian noise. Since the detection of *bandpass* signals employs the same concepts, we shall summarize the key findings of that section. We focus particularly on that realization of a matched filter known as a *correlator.* In addition to binary detection, we also consider the more general case of M-ary detection. We assume that the only performance degradation is due to AWGN. The received signal is the sum of the transmitted prototype signal plus the random noise:

$$r(t) = s_i(t) + n(t) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.14}$$

Given such a received signal, the detection process consists of *two basic steps* as shown in Figure 3.1. In the first step, the received waveform, $r(t)$, is reduced to a *single random variable* $z(T)$, or to a *set of random variables* $z_i(T)$ $(i = 1, \ldots, M)$, formed at the output of the demodulator and sampler at time $t = T$, where T is the symbol duration. In the second step, a symbol decision is made on the basis of comparing $z(T)$ to a threshold or on the basis of choosing the maximum $z_i(T)$. Step 1 can be thought of as transforming the waveform into a point in the decision space. This point can be referred to as the *predetection point,* the most critical reference point in the receiver. When we talk about received signal power, or received interfering noise, or E_b/N_0, their values are always considered with reference to this predetection point. Sometimes such received parameters are loosely described with reference to the *receiver input* or the *receiving antenna.* But when described in such ways, there is an underlying assumption of no loss in SNR or E_b/N_0 between these inputs and the predetection point. At each symbol time, the signal that is available at the predetection point is a sample of a baseband pulse. We do not yet have bits.

Does the fact that the energy-per-bit versus N_0 is *defined* at a place where there are no bits strike the reader as a paradox? Actually, it really should not, because this is a convenient reference point, where the baseband pulse—even before bit-decisions are made can be considered to *effectively* represent bits. Step 2 can be thought of as determining *in which decision region* the point is located. For the detector to be optimized (in the sense of minimizing the error probability), it is necessary to optimize the waveform-to-random-variable transformation, by using matched filters or correlators in step 1, and by also optimizing the decision criterion in step 2.

In Sections 3.2.2 and 3.2.3 we found that the matched filter provides the maximum signal-to-noise ratio at the filter output at time $t = T$. We described a correlator as one realization of a matched filter. We can define a *correlation receiver* comprised of M correlators, as shown in Figure 4.7a, that transforms a received waveform, $r(t)$, to a sequence of M numbers or correlator outputs, $z_i(T)$ $(i = 1, \ldots, M)$. Each correlator output is characterized by the following product integration or correlation with the received signal:

$$z_i(T) = \int_0^T r(t)s_i(t)\,dt \qquad i = 1, \ldots, M \tag{4.15}$$

The verb "to correlate" means "to match." The correlators attempt to match the incoming received signal, $r(t)$, with each of the candidate prototype waveforms, $s_i(t)$, known a priori to the receiver. A reasonable decision rule is to choose the waveform, $s_i(t)$, that *matches best* or has the *largest correlation* with $r(t)$. In other words, the decision rule is

$$\begin{array}{l}\text{Choose the } s_i(t) \text{ whose index} \\ \text{corresponds to the max } z_i(T)\end{array} \tag{4.16}$$

Following Equation (3.10), any signal set, $\{s_i(t)\}$ $(i = 1, \ldots, M)$, can be expressed in terms of some set of basis functions, $\{\psi_j(t)\}$ $(j = 1, \ldots, N)$, where $N \leq M$. Then the bank of M correlators in Figure 4.7a may be replaced with a bank of N correlators, shown in Figure 4.7b, where the set of basis functions $\{\psi_j(t)\}$ form *reference signals.* The decision stage of this receiver consists of logic circuitry for choosing the signal, $s_i(t)$. The choice of $s_i(t)$ is made according to the best match of the coefficients, a_{ij}, seen in Equation (3.10), with the set of outputs $\{z_j(T)\}$. When the prototype waveform set, $\{s_i(t)\}$, is an orthogonal set, the receiver implementation in Figure 4.7a is identical to that in Figure 4.7b (differing perhaps by a scale factor). However, when $\{s_i(t)\}$ is *not* an orthogonal set, the receiver in Figure 4.7b, using N correlators instead of M, with reference signals $\{\psi_j(t)\}$, can represent a cost-effective implementation. We examine such an application for the detection of multiple phase shift keying (MPSK) in Section 4.4.3.

In the case of *binary detection,* the correlation receiver can be configured as a single matched filter or product integrator, as shown in Figure 4.8a, with the reference signal being the difference between the binary prototype signals, $s_1(t) - s_2(t)$. The output of the correlator, $z(T)$, is fed directly to the decision stage.

For binary detection, the correlation receiver can also be drawn as two matched filters or product integrators, one of which is matched to $s_1(t)$, and the

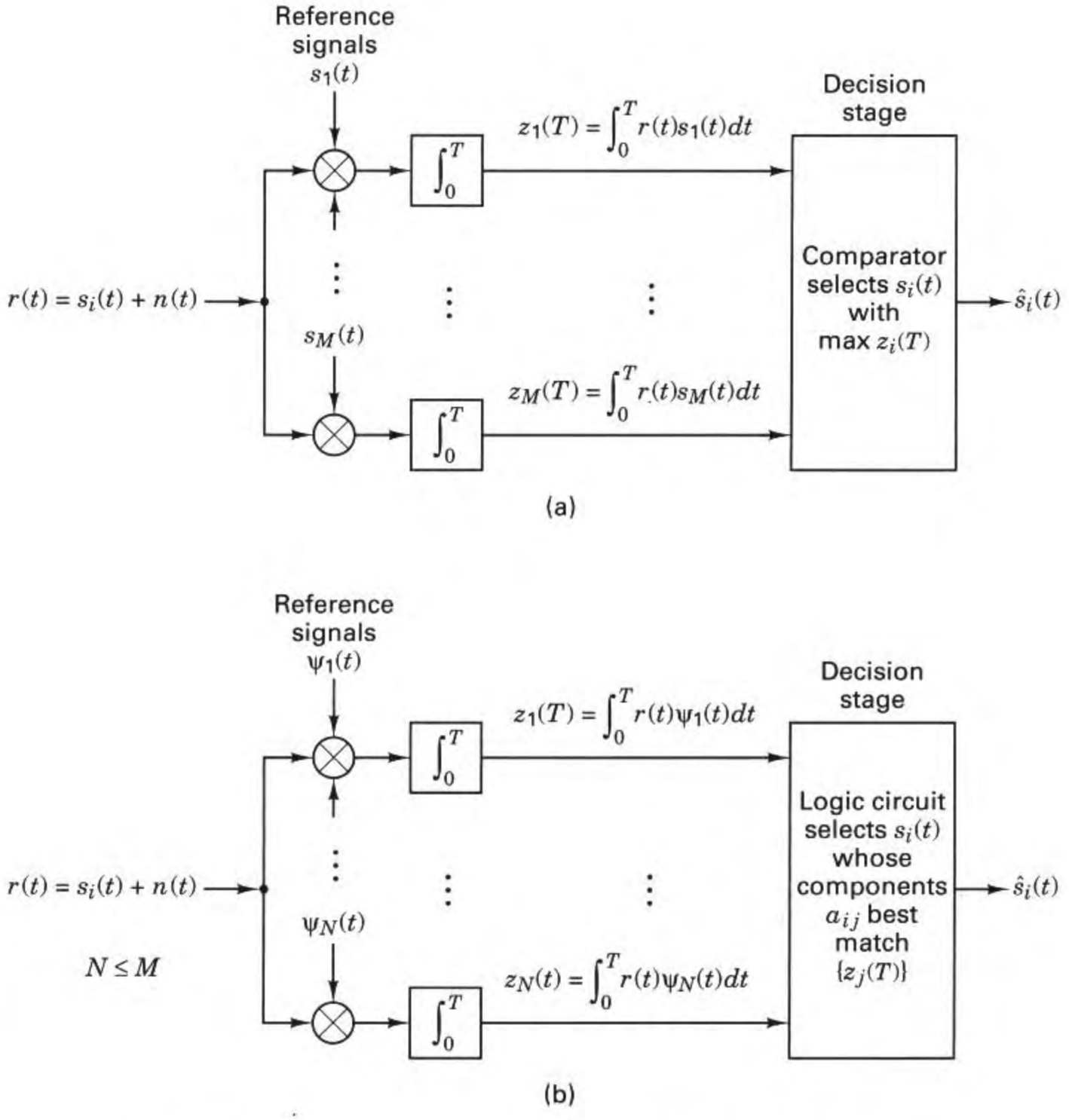


Figure 4.7 (a) Correlator receiver with reference signals $\{s_i(t)\}$. (b) Correlator receiver with reference signals $\{\psi_j(t)\}$.

other is matched to $s_2(t)$. (See Figure 4.8b.) The decision stage can then be configured to follow the rule in Equation 4.16, or the correlator outputs $z_i(T)$ $(i = 1, 2)$ can be differenced to form

$$z(T) = z_1(T) - z_2(T) \tag{4.17}$$

as shown in Figure 4.8b. Then, $z(T)$, called the *test statistic,* is fed to the decision stage, as in the case of the single correlator. In the *absence of noise,* an input waveform $s_i(t)$ yields the output $z(T) = a_i(T)$, a signal-only component. The input noise

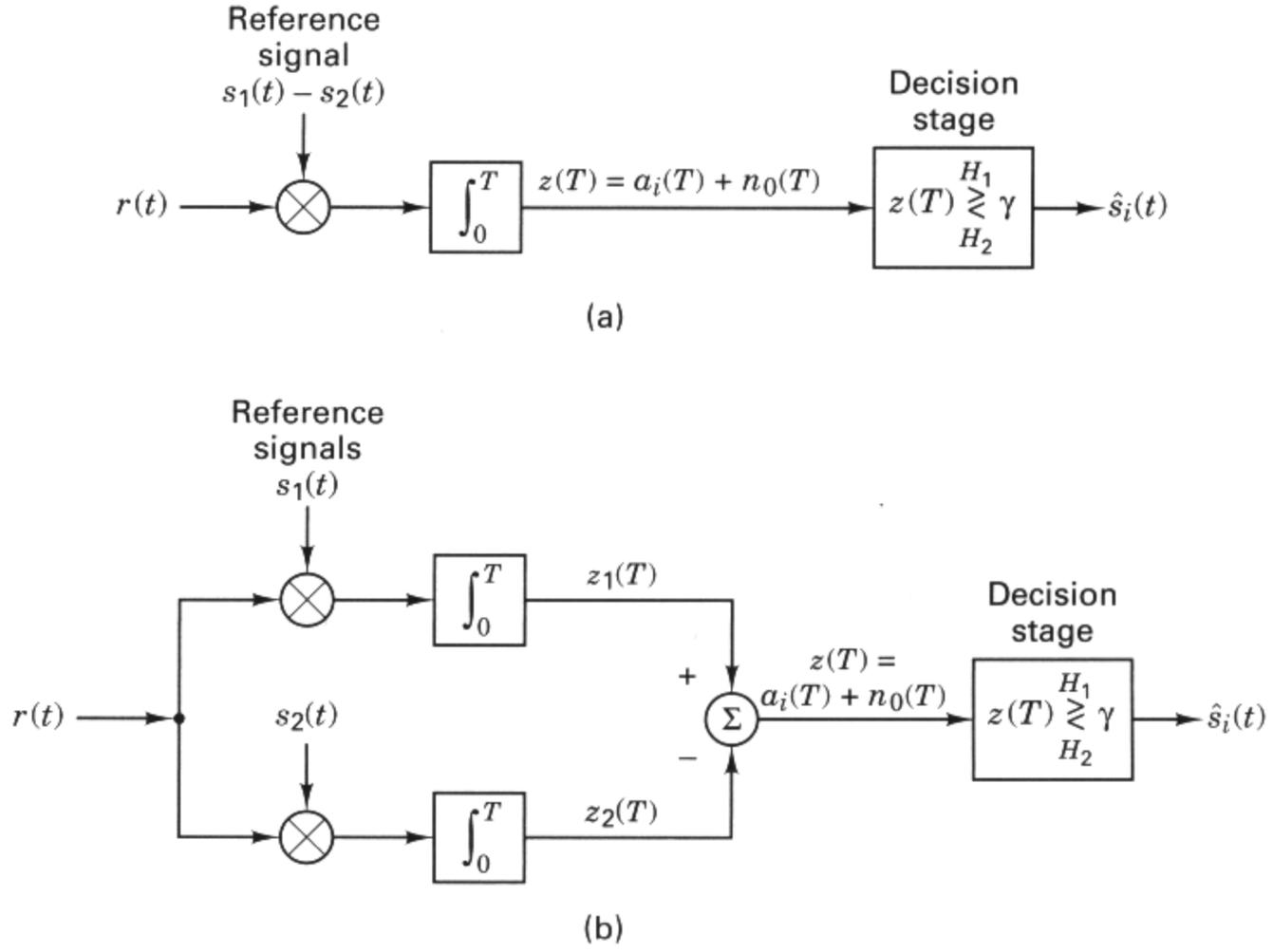


Figure 4.8 Binary correlator receiver. (a) Using a single correlator. (b) Using two correlators.

$n(T)$ is a Gaussian random process. Since the correlator is a *linear* device, the output noise is also a Gaussian random process [2]. Thus, the output of the correlator, sampled at $t = T$, yields

$$z(T) = a_i(T) + n_0(T) \qquad i = 1, 2$$

where $n_0(T)$ is the noise component. To shorten the notation we sometimes express $z(t)$ as $a_i + n_0$. The noise component n_0 is a zero-mean *Gaussian random variable,* and thus $z(T)$ is a *Gaussian random variable* with a mean of either a_1 or a_2, depending on whether a binary one or binary zero was sent.

4.3.2.1 Binary Decision Threshold

For the random variable $z(T)$, Figure 4.9 illustrates the two conditional probability density functions (pdfs), $p(z|s_1)$ and $p(z|s_2)$, with mean value of a_1 and a_2, respectively. These pdfs, also called the *likelihood* of s_1 and the *likelihood* of s_2, respectively, were presented in Section 3.1.2, and are rewritten as

$$p(z|s_1) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z - a_1}{\sigma_0}\right)^2\right] \tag{4.18a}$$

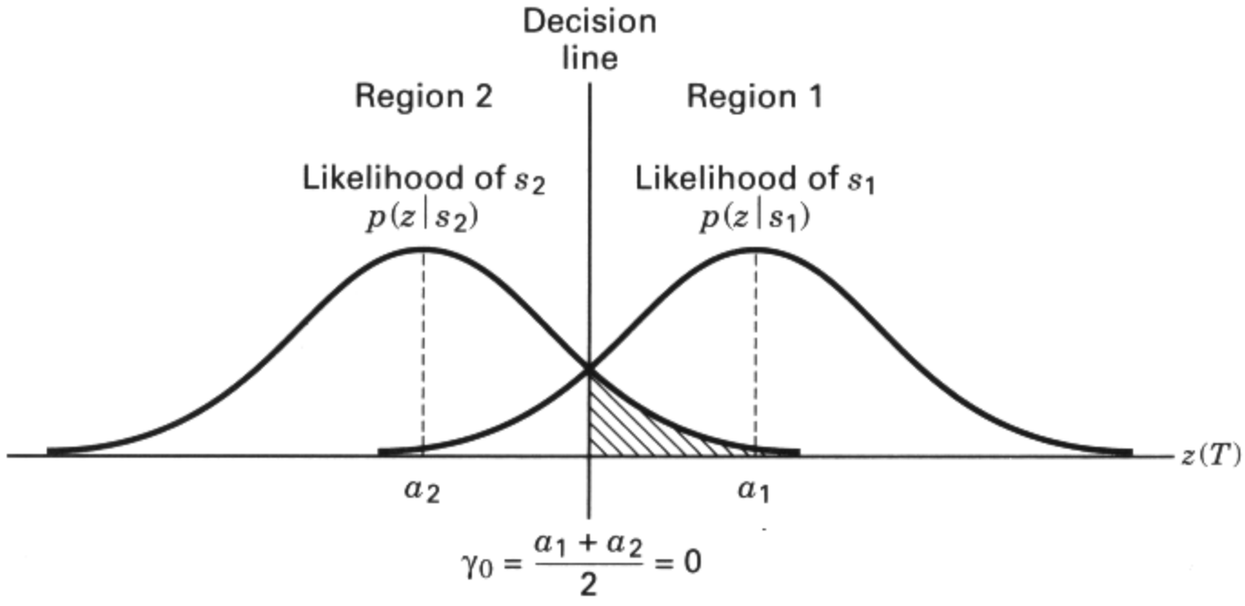


Figure 4.9 Conditional probability density functions: $p(z/s_1)$, $p(z/s_2)$.

and

$$p(z|s_2) = \frac{1}{\sigma_0\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{z - a_2}{\sigma_0}\right)^2\right] \tag{4.18b}$$

where σ_0^2 is the noise variance. In Figure 4.9 the rightmost likelihood $p(z|s_1)$ illustrates the probability density of the detector output $z(T)$, given that $s_1(t)$ was transmitted. Similarly, the leftmost likelihood $p(z|s_2)$ illustrates the probability density of $z(T)$, given that $s_2(t)$ was transmitted. The abscissa $z(T)$ represents the full range of possible sample output values from the correlation receiver shown in Figure 4.8.

With regard to optimizing the binary decision threshold for deciding in which region a received signal is located, we found in Section 3.2.1 that the *minimum error* criterion for equally likely binary signals corrupted by Gaussian noise can be stated as

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} \frac{a_1 + a_2}{2} = \gamma_0 \tag{4.19}$$

where a_1 is the signal component of $z(T)$ when $s_1(t)$ is transmitted, and a_2 is the signal component of $z(T)$ when $s_2(t)$ is transmitted. The threshold level γ_0 represented by $(a_1 + a_2)/2$ is the *optimum threshold* for minimizing the probability of making an incorrect decision given equally likely signals and symmetrical likelihoods. The decision rule in Equation (4.19) states that hypothesis H_1 should be selected [equivalent to deciding that signal $s_1(t)$ was sent] if $z(T) > \gamma_0$, and hypothesis H_2 should be selected [equivalent to deciding that $s_2(t)$ was sent] if $z(T) < \gamma_0$. If $z(T) = \gamma_0$, the decision can be an arbitrary one. For equal-energy, equally likely antipodal signals, where $s_1(t) = -s_2(t)$ and $a_1 = -a_2$, the optimum decision rule becomes

$$z(T) \underset{H_2}{\overset{H_1}{\gtrless}} = 0_1 \tag{4.20a}$$

or

$$\begin{array}{ll} \text{decide } s_1(t) & \text{if } z_1(T) > z_2(T) \\ \text{decide } s_2(t) & \text{otherwise} \end{array} \tag{4.20b}$$

4.4 COHERENT DETECTION

4.4.1 Coherent Detection of PSK

The detector shown in Figure 4.7 can be used for the coherent detection of any digital waveforms. Such a correlating detector is often referred to as a *maximum likelihood detector.* Consider the following binary PSK (BPSK) example: Let

$$s_1(t) = \sqrt{\frac{2E}{T}} \cos(\omega_0 t + \phi) \qquad 0 \le t \le T \tag{4.21a}$$

$$\begin{aligned} s_2(t) &= \sqrt{\frac{2E}{T}} \cos(\omega_0 t + \phi + \pi) \\ &= -\sqrt{\frac{2E}{T}} \cos(\omega_0 t + \phi) \qquad 0 \le t \le T \end{aligned} \tag{4.21b}$$

and

$$n(t) = \text{zero-mean white Gaussian random process}$$

where the phase term ϕ is an arbitrary constant, so that the analysis is unaffected by setting $\phi = 0$. The parameter E is the signal energy per symbol, and T is the symbol duration. For this antipodal case, only a single basis function is needed. If an orthonormal signal space is assumed in Equations (3.10) and (3.11) (i.e., $K_j = 1$), we can express a basis function $\psi_1(t)$ as

$$\psi_1(t) = \sqrt{\frac{2}{T}} \cos \omega_0 t \qquad \text{for } 0 \le t \le T \tag{4.22}$$

Thus, we may express the transmitted signals $s_i(t)$ in terms of $\psi_1(t)$ and the coefficients $a_{i1}(t)$ as follows:

$$s_i(t) = a_{i1}\psi_1(t) \tag{4.23a}$$

$$s_1(t) = a_{11}\psi_1(t) = \sqrt{E}\,\psi_1(t) \tag{4.23b}$$

$$s_2(t) = a_{21}\psi_1(t) = -\sqrt{E}\,\psi_1(t) \tag{4.23c}$$

Assume that $s_1(t)$ was transmitted. Then the expected values of the product integrators in Figure 4.7b, with reference signal $\psi_1(t)$, are found as

$$\mathbf{E}\{z_1|s_1\} = \mathbf{E}\left\{ \int_0^T \sqrt{E}\,\psi_1^2(t) + n(t)\psi_1(t)\,dt \right\} \tag{4.24a}$$

$$\mathbf{E}\{z_2|s_1\} = \mathbf{E}\left\{ \int_0^T -\sqrt{E}\,\psi_1^2(t) + n(t)\psi_1(t)\,dt \right\} \tag{4.24b}$$

$$\mathbf{E}\{z_1|s_1\} = \mathbf{E}\left\{ \int_0^T \frac{2}{T}\sqrt{E}\cos^2\omega_0 t + n(t)\sqrt{\frac{2}{T}}\cos\omega_0 t\,dt \right\} = \sqrt{E} \tag{4.25a}$$

and

$$\mathbf{E}\{z_2|s_1\} = \mathbf{E}\left\{ \int_0^T -\frac{2}{T}\sqrt{E}\cos^2\omega_0 t + n(t)\sqrt{\frac{2}{T}}\cos\omega_0 t\,dt \right\} = -\sqrt{E} \tag{4.25b}$$

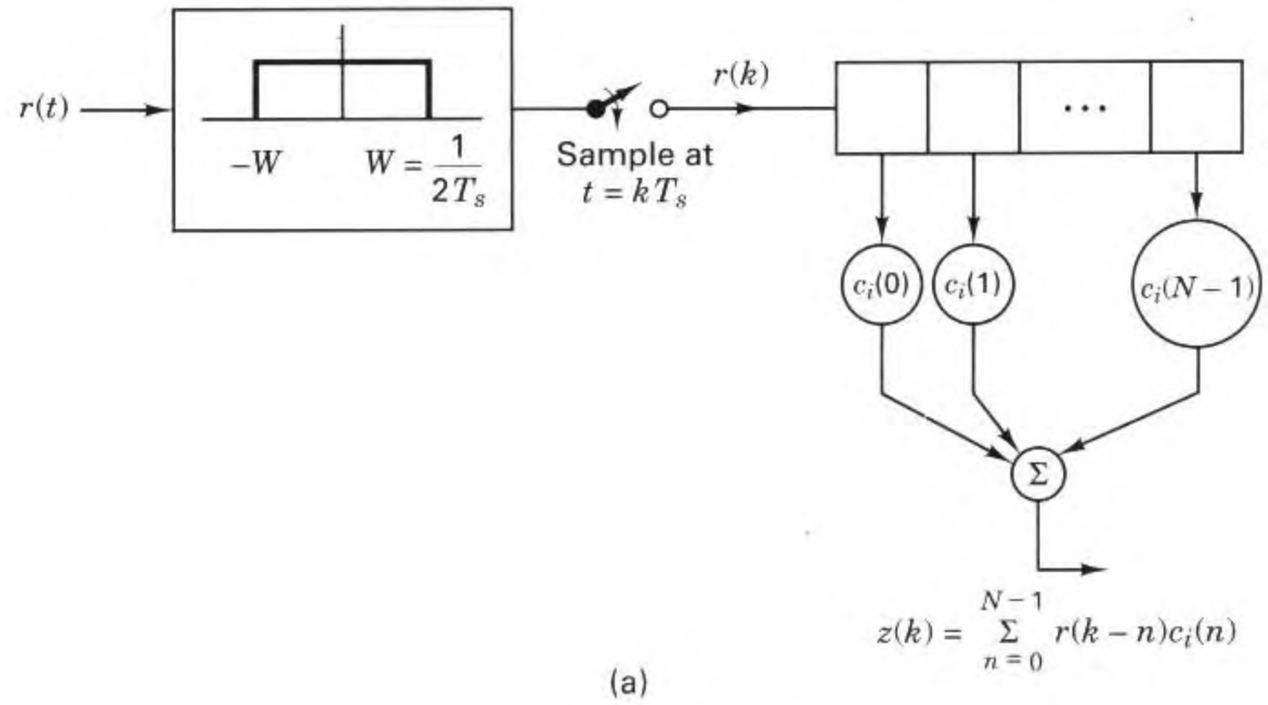
where E{·} denotes the ensemble average, referred to as the *expected value.* Equation (4.25) follows because $\mathbf{E}\{n(t)\} = 0$. The decision stage must decide which signal was transmitted by determining its location within the signal space. For this example, the choice of $\psi_1(t) = \sqrt{2/T}\cos\omega_0 t$ normalizes $\mathbf{E}\{z_i(T)\}$ to be $\pm\sqrt{E}$. The prototype signals $\{s_i(t)\}$ are the same as the reference signals $\{\psi_j(t)\}$ except for the normalizing scale factor. The decision stage chooses the signal with the largest value of $z_i(T)$. Thus, the received signal in this example is judged to be $s_1(t)$. The error performance for such coherently detected BPSK systems is treated in Section 4.7.1.

4.4.2 Sampled Matched Filter

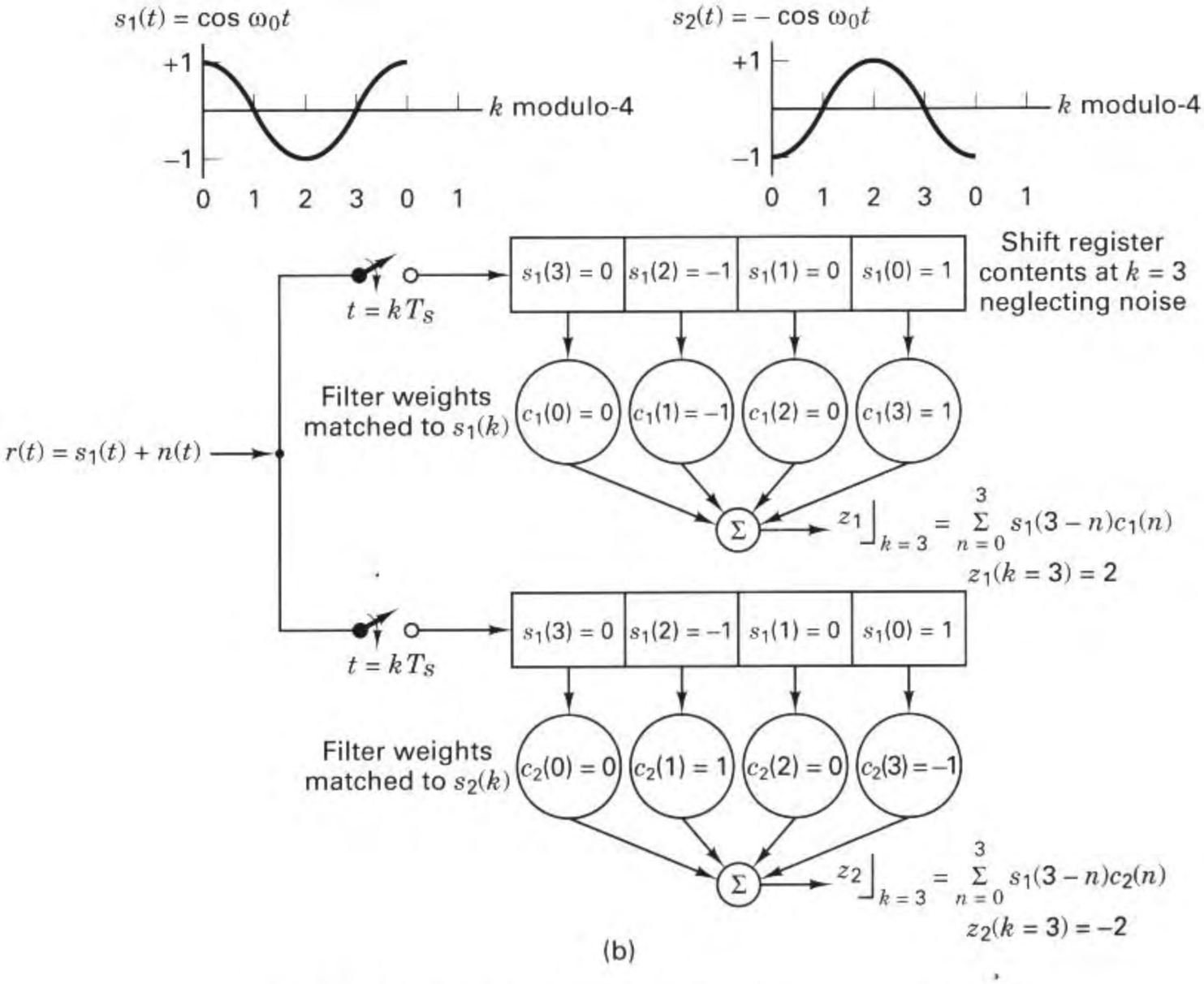
In Section 3.2.2, we discussed the basic characteristic of the matched filter—namely, that its impulse response is a delayed version of the mirror image (rotated on the $t = 0$ axis) of the input signal waveform. Therefore, if the signal waveform is $s(t)$, its mirror image is $s(-t)$, and the mirror image delayed by T seconds is $s(T-t)$. The impulse response $h(t)$ of a filter matched to $s(t)$ is then described by

$$h(t) = \begin{cases} s(T-t) & 0 \le t \le T \\ 0 & \text{elsewhere} \end{cases} \tag{4.26}$$

Figures 4.7 and 4.8 illustrate the basic function of a correlator to product-integrate the received noisy signal with each of the candidate reference signals and determine the best match. The schematics in these figures imply the use of analog hardware (multipliers and integrators) and continuous signals. They do not reflect the way that the correlator or matched filter (MF) can be implemented using digital techniques and sampled waveforms. Figure 4.10 shows how an MF can be implemented using digital hardware. The input signal $r(t)$ comprises a prototype signal $s_i(t)$, plus noise $n(t)$, and the bandwidth of the signal is $W = 1/2T$, where T is the symbol time. Thus, the minimum Nyquist sampling rate is $f_s = 2W = 1/T$, and the sampling time T_s needs to be equal to or less than the symbol time. In other words, there must be at least one sample per symbol. In real systems, such sampling is usually performed at a rate that exceeds the Nyquist minimum by a factor of 4 or more. The only cost is processor speed, not transmission bandwidth. At the clock



(a)



(b)

Figure 4.10 (a) Sampled matched filter. (b) Sampled matched filter detection example, neglecting noise.

times of $t = kT_s$, the samples are shifted into the register (as shown in Figure 4.10a) so that earlier samples are located to the right of later samples. Once the received signal has been sampled, the continuous time notation t is changed to kT_s or simply to k to allow use of the simple discrete notation

$$r(k) = s_i(k) + n(k) \qquad i = 1, 2 \qquad k = 0, 1, \ldots$$

where the index i identifies a particular symbol out of the M-ary set (binary in this example), and k is the sampling-time index. In Figure 4.10, the shift register, with its coefficients or weights $c_i(n)$ approximate an MF, where $n = 0, \ldots, N - 1$ is the time index of the weights and register stages. In this example, $N = 4$ represents the number of stages in the register and the number of samples per symbol. Hence, the summation shown in the figure takes place over the time $n = 0$ to $n = 3$. Following the summer, a symbol decision will be made after 4 time samples have entered the registers. Note that for simplicity, the example in Figure 4.10b uses only 3-level amplitude values (0, ±1) for the $s_i(k)$ samples. In real systems, each sample (and weight) would be represented by 6–10 bits, depending on the density of the signal constellation. The set of filter weights $\{c_i(n)\}$ constitutes the filter impulse response; the weights are matched to signal samples according to the discrete form of Equation (4.26), as follows:

$$c_i(n) = s_i[(N - 1) - n] = s_i(3 - n) \tag{4.27}$$

By using the discrete form of the *convolution integral* in Equation (A.44b), the output at a time corresponding to the kth sample can be expressed as

$$z_i(k) = \sum_{n=0}^{N-1} r(k - n)\, c_i(n) \qquad k = 0, 1, \ldots, \text{modulo-}N \tag{4.28}$$

where x modulo-y is defined as the remainder of dividing x by y, the index k represents time for both the received samples and the filter output, and n is a dummy time variable. In the expression $r(k - n)$ of Equation (4.28), it is helpful to think of n as the "age" of the sample (how long has it been sitting in the filter). In the expression $c_i(n)$, it is helpful to think of n as the address of the weight. We assume that because the system is synchronized, the symbol timing is known. Also, we assume the noise to have a zero mean, so that the expected value of a received sample is

$$\mathbf{E}\{r(k)\} = s_i(k) \qquad i = 1, 2$$

Thus, if $s_i(t)$ is transmitted, the expected matched filter output is

$$\mathbf{E}\{z_i(k)\} = \sum_{n=0}^{N-1} s_i(k - n)\, c_i(n) \qquad k = 0, 1, \ldots, \text{modulo-}N \tag{4.29}$$

In Figure 4.10b, where the prototype waveforms are plotted as a function of time, we see that the leftmost sample (amplitude equals + 1) of the $s_1(t)$ plot represents the earliest sample at time $k = 0$. Assuming that $s_1(t)$ was sent, and the noise is neglected for notational simplicity, we then denote the received samples of $r(k)$ as $s_1(k)$. As these samples fill the stages of the MF, at the end of each symbol time the

$k = 0$ sample is located in the rightmost stage of each register. In Equations (4.28) and (4.29), notice that the time indexes n of the reference weights are in reverse order compared with the time index $k - n$ of the samples, which is a key aspect of the convolution integral. The fact that the earliest time sample now corresponds to the rightmost weight will ensure a meaningful correlation. Even though we describe the mathematical operation of an MF to be *convolution* of a signal with the impulse response of the filter, the end result appears to be the *correlation* of a signal with a replica of that same signal. That is why it is valid to describe a correlator as an implementation of a matched filter.

In Figure 4.10b, detection will follow the MF in the usual way. For the binary decision, the $z_i(k)$ outputs are examined at each value of $k = N - 1$ corresponding to the end of a symbol. Under the condition that $s_1(t)$ had been transmitted and noise is neglected, we combine Equations (4.27) through (4.29) to express the correlator outputs at time $k = N - 1 = 3$ as

$$z_1(k=3) = \sum_{n=0}^{3} s_1(3 - n)\, c_1(n) = 2 \tag{4.30a}$$

and

$$z_2(k=3) = \sum_{n=0}^{3} s_1(3 - n)\, c_2(n) = -2 \tag{4.30b}$$

Since $z_1(k = 3)$ is greater than $z_2(k = 3)$, the detector chooses $s_1(t)$ as the transmitted symbol.

One might ask, "What is the difference between the MF in Figure 4.10b and the correlator in Figure 4.8?" In the case of the MF, a new output value is available in response to each new input sample; thus the output will be a time series such as the MF output seen in Figure 3.7b (a succession of increasing positive and negative correlations to an input sine wave). Such an MF output sequence can be equated to several correlators operating at different starting points of the input time series. Note that a correlator only computes an output once per symbol time, such as the value of the peak signal at time T in Figure 3.7b. If the timing of the MF and correlator are aligned, then their outputs at the end of a symbol time are identical. An important distinction between the MF and correlator is that since the correlator yields a single output value per symbol, it must have side information, such as the start and stop times over which the product integration should take place. If there are timing errors in the correlator, then the sampled output fed to the detector may be badly degraded. On the other hand, since the MF yields a *time series* of output values (reflecting time-shifted input samples multiplied by fixed weights), then with the use of additional circuitry, the best time for sampling the MF output can be learned.

Example 4.1 Sampled Matched Filter

Consider the waveform set

$$s_1(t) = At \qquad 0 \le t \le kT$$

and

$$s_2(t) = -At \qquad 0 \le t \le kT$$

where $k = 0, 1, 2, 3$.

Illustrate how a *sampled* matched filter as shown in Figure 4.10 can be used to detect a received signal, say $s_1(t)$, from this sawtooth waveform set in the absence of noise.

Solution

First, the waveform is sampled so that $s_1(t)$ is transformed into the set of samples $\{s_1(k)\}$. The sampled matched filter receiver will be shown with two branches, following the implementation in Figure 4.10b. The top branch is made up of shift registers and coefficients matched to the $\{s_1(k)\}$ sample points. The bottom branch is similarly matched to the $\{s_2(k)\}$ sample points. The four equally spaced sample points ($k = 0, 1, 2, 3$) for each of the $\{s_i(k)\}$ are as follows:

$$s_1(k=0) = 0 \qquad s_1(k=1) = \frac{A}{4} \qquad s_1(k=2) = \frac{A}{2} \qquad s_1(k=3) = \frac{3A}{4}$$

$$s_2(k=0) = 0 \qquad s_2(k=1) = -\frac{A}{4} \qquad s_2(k=2) = -\frac{A}{2} \qquad s_2(k=3) = -\frac{3A}{4}$$

The $c_i(n)$ coefficients represent the delayed mirror-image rotation of the signal to which the filter is matched. Therefore, $c_i(n) = s_i(N-1-n)$, where $n = 0, \ldots, N-1$, and we can write $c_i(0) = s_i(3)$, $c_i(1) = s_i(2)$, $c_i(2) = s_i(1)$, $c_i(3) = s_i(0)$.

Consider the top branch in Figure 4.10b. At the $k = 0$ clock time, the first sample $s_1(k=0) = 0$ enters the leftmost stage of each register. At the next clock time, the second sample $s_1(k=1) = A/4$ enters the leftmost stage of each register; at this same time the first sample has been shifted to the next right stage in each register, and so on. At the $k = 3$ clock time the sample $s_1(k=3) = 3A/4$ enters the leftmost stage; by this time the first sample has been shifted into the rightmost stage. The four signal samples are now located in the registers in mirror-image arrangement compared with how they would be plotted in time. Hence, the convolution operation is an appropriate expression for describing the alignment of the incoming waveform samples with the reference coefficients to maximize the correlation in the proper branch.

4.4.3 Coherent Detection of Multiple Phase-Shift Keying

Figure 4.11 illustrates the signal space for a multiple phase-shift keying (MPSK) signal set; the figure describes a four-level (4-ary) PSK or quadriphase shift keying (QPSK) example ($M = 4$). At the transmitter, binary digits are collected two at a time, and for each symbol interval, the two sequential digits instruct the modulator as to which of the four waveforms to produce. For typical coherent M-ary PSK (MPSK) systems, $s_i(t)$ can be expressed as

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos\left(\omega_0 t - \frac{2\pi i}{M}\right) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.31}$$

where E is the received energy of such a waveform over each symbol duration T, and ω_0 is the carrier frequency. If an orthonormal signal space is assumed in Equations (3.10) and (3.11), we can choose a convenient set of axes, such as

$$\psi_1(t) = \sqrt{\frac{2}{T}} \cos \omega_0 t \tag{4.32a}$$

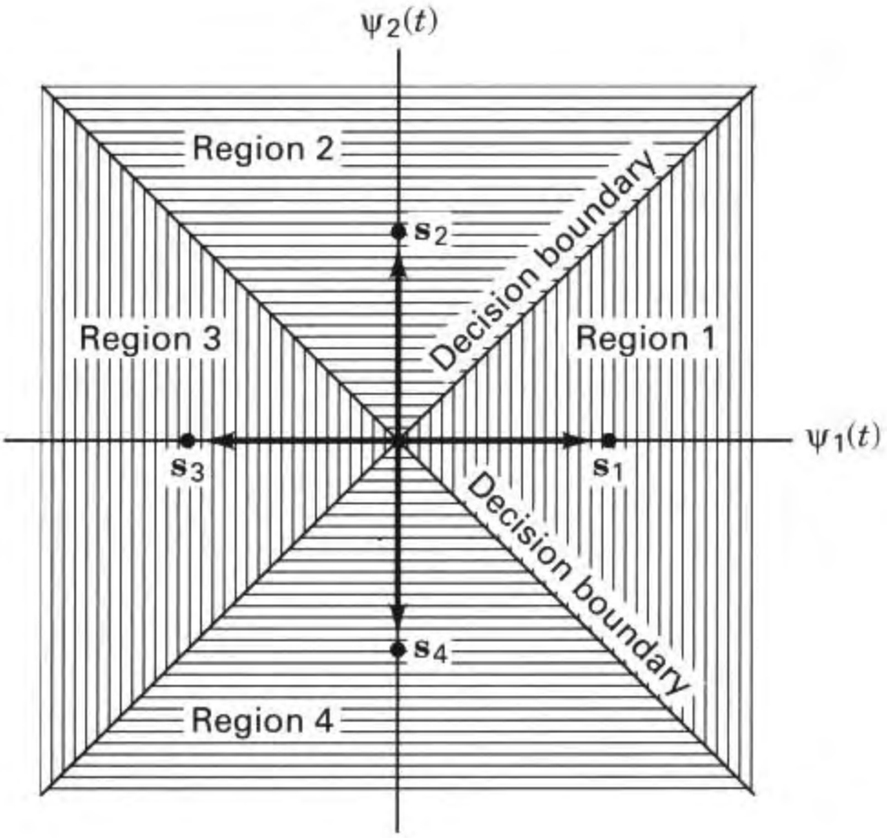


Figure 4.11 Signal space and decision regions for a QPSK system.

and

$$\psi_2(t) = \sqrt{\frac{2}{T}} \sin \omega_0 t \tag{4.32b}$$

where the amplitude $\sqrt{2/T}$ has been chosen to normalize the expected output of the detector, as was done in Section 4.4.1. Now $s_i(t)$ can be written in terms of these orthonormal coordinates, giving

$$s_i(t) = a_{i1}\psi_1(t) + a_{i2}\psi_2(t) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.33a}$$

$$= \sqrt{E} \cos\left(\frac{2\pi i}{M}\right) \psi_1(t) + \sqrt{E} \sin\left(\frac{2\pi i}{M}\right) \psi_2(t) \tag{4.33b}$$

Notice that Equation (4.33) describes a set of M multiple phase waveforms (intrinsically nonorthogonal) in terms of only two orthogonal carrier-wave components. The $M = 4$ (QPSK) case is unique among MPSK signal sets in the sense that the QPSK waveform set is represented by a combination of antipodal and orthogonal members. The decision boundaries partition the signal space into $M = 4$ regions; the construction is similar to the procedure outlined in Section 4.3.1 and Figure 4.6 for $M = 2$. The decision rule for the detector (see Figure 4.11) is to decide that $s_1(t)$ was transmitted if the received signal vector falls in region 1, that $s_2(t)$ was transmitted if the received signal vector falls in region 2, and so on. In other words, the decision rule is to choose the ith waveform if $z_i(T)$ is the largest of the correlator outputs (seen in Figure 4.7).

The form of the correlator shown in Figure 4.7a implies that there are always M product correlators used for the demodulation of MPSK signals. The figure infers that for each of the M branches, a reference signal with the appropriate phase shift is configured. In practice, the implementation of an MPSK demodulator follows Figure 4.7b, requiring only $N = 2$ product integrators regardless of the size of the signal set M. The savings in implementation is possible because any arbitrary integrable waveform set can be expressed as a linear combination of orthogonal waveforms, as shown in Section 3.1.3. Figure 4.12 illustrates such a demodulator. The received signal $r(t)$ can be expressed by combining Equations (4.32) and (4.33) as

$$r(t) = \sqrt{\frac{2E}{T}}(\cos\phi_i \cos\omega_0 t + \sin\phi_i \sin\omega_0 t) + n(t) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.34}$$

where $\phi_i = 2\pi i/M$, and $n(t)$ is a zero-mean white Gaussian noise process. Notice that in Figure 4.12, there are only two reference waveforms or basis functions, $\psi_1(t) = \sqrt{2/T}\cos\omega_0 t$ for the upper correlator and $\psi_2(t) = \sqrt{2/T}\sin\omega_0 t$ for the lower correlator. The upper correlator computes

$$X = \int_0^T r(t)\psi_1(t)\,dt \tag{4.35}$$

and the lower correlator computes

$$Y = \int_0^T r(t)\psi_2(t)\,dt \tag{4.36}$$

Figure 4.13 illustrates that the computation of the received phase angle $\hat{\phi}$ can be accomplished by computing the arctan of Y/X, where X can be thought of as the in-phase component of the received signal, Y is the quadrature component, and $\hat{\phi}$ is a

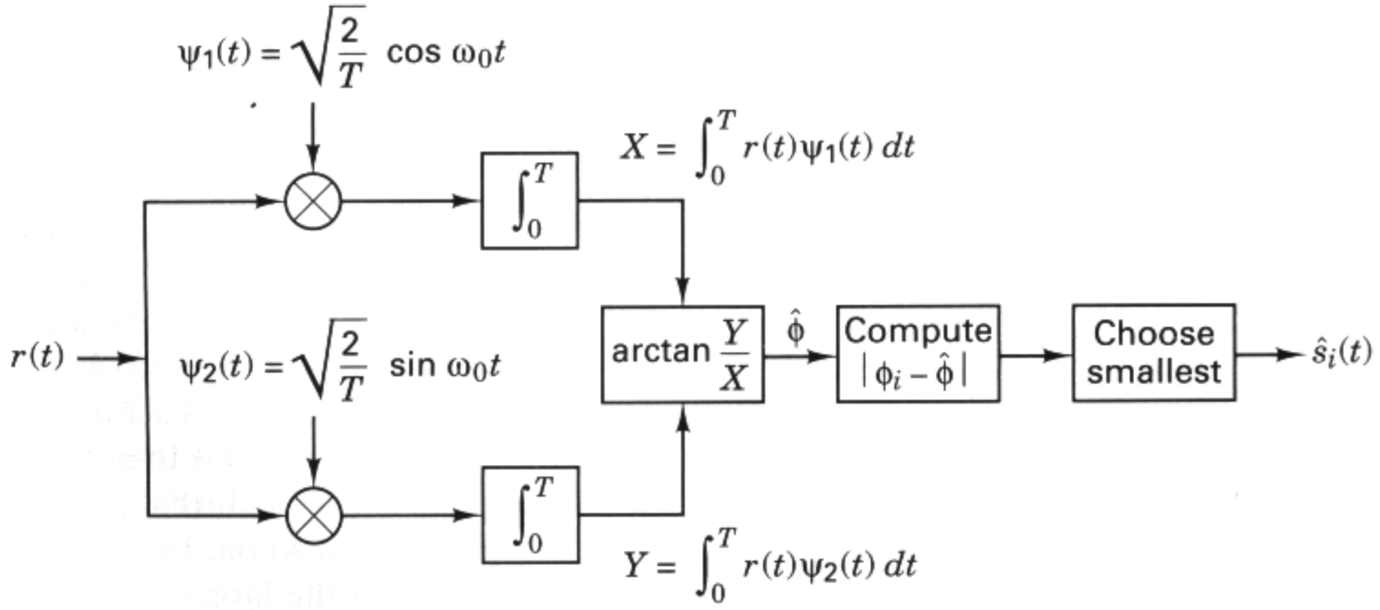


Figure 4.12 Demodulator for MPSK signals.

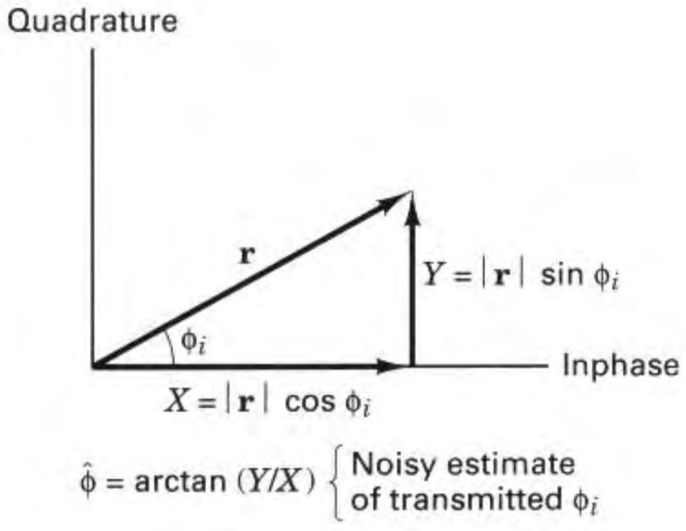


Figure 4.13 In-phase and quadrature components of the received signal vector **r**.

noisy estimate of the transmitted ϕ_i. In other words, the upper correlator of Figure 4.12 produces an output X, the magnitude of the in-phase projection of the vector **r**, and the lower correlator produces an output Y, the magnitude of the quadrature projection of the vector **r**, where **r** is the vector representation of $r(t)$. The X and Y outputs of the correlators feed into the block marked arctan (Y/X). The resulting value of the angle $\hat{\phi}$ is compared with each of the prototype phase angles, ϕ_i. The demodulator selects the ϕ_i that is closest to the angle $\hat{\phi}$. In other words, the demodulator computes $|\phi_i - \hat{\phi}|$ for each of the ϕ_i prototypes and chooses the ϕ_i yielding the smallest output.

4.4.4 Coherent Detection of FSK

FSK modulation is characterized by the information being contained in the frequency of the carrier. A typical set of FSK signal waveforms was described in Equation (4.8) as

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix}$$

where E is the energy content of $s_i(t)$ over each symbol duration T, and $(\omega_{i+1} - \omega_i)$ is typically assumed to be an integral multiple of π/T. The phase term ϕ is an arbitrary constant and can be set equal to zero. Assuming that the basis functions $\psi_1(t)$, $\psi_2(t), \ldots, \psi_N(t)$ form an orthonormal set, the most useful form for $\{\psi_j(t)\}$ is

$$\psi_j(t) = \sqrt{\frac{2}{T}} \cos \omega_j t \qquad j = 1, \ldots, N \tag{4.37}$$

where, as before, the amplitude $\sqrt{2/T}$ normalizes the expected output of the matched filter. From Equation (3.11), we can write

$$a_{ij} = \int_0^T \sqrt{\frac{2E}{T}} \cos(\omega_i t) \sqrt{\frac{2}{T}} \cos \omega_j t \, dt \tag{4.38}$$

Therefore,

$$a_{ij} = \begin{cases} \sqrt{E} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{4.39}$$

In other words, the ith prototype signal vector is located on the ith coordinate axis at a displacement $\sqrt{E}$ from the origin of the signal space. In this scheme, for the general M-ary case and a given E, the distance between any two prototype signal vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ is constant:

$$d(\mathbf{s}_i, \mathbf{s}_j) = \|\mathbf{s}_i - \mathbf{s}_j\| = \sqrt{2E} \qquad \text{for } i \neq j \tag{4.40}$$

Figure 4.14 illustrates the prototype signal vectors and the decision regions for a 3-ary ($M = 3$) coherently detected orthogonal FSK signaling scheme. A natural choice for the size M of a signaling set is any power-of-two value. However in this case, the reasons for the unorthodox selection of $M = 3$, are: We desire to examine a signaling set that is greater than binary, and the signaling space for orthogonal signaling is best visualized as having mutually perpendicular axes. It is not possible, beyond 3-axes, to convey the notion of mutual perpendicularity in some visually accurate way. As in the PSK case, the signal space is partitioned into M distinct regions, each containing one prototype signal vector; in this example, because the decision region is three-dimensional, the decision boundaries are planes instead of lines. The optimum decision rule is to decide that the transmitted signal belongs to the class whose index corresponds to the region where the received signal is

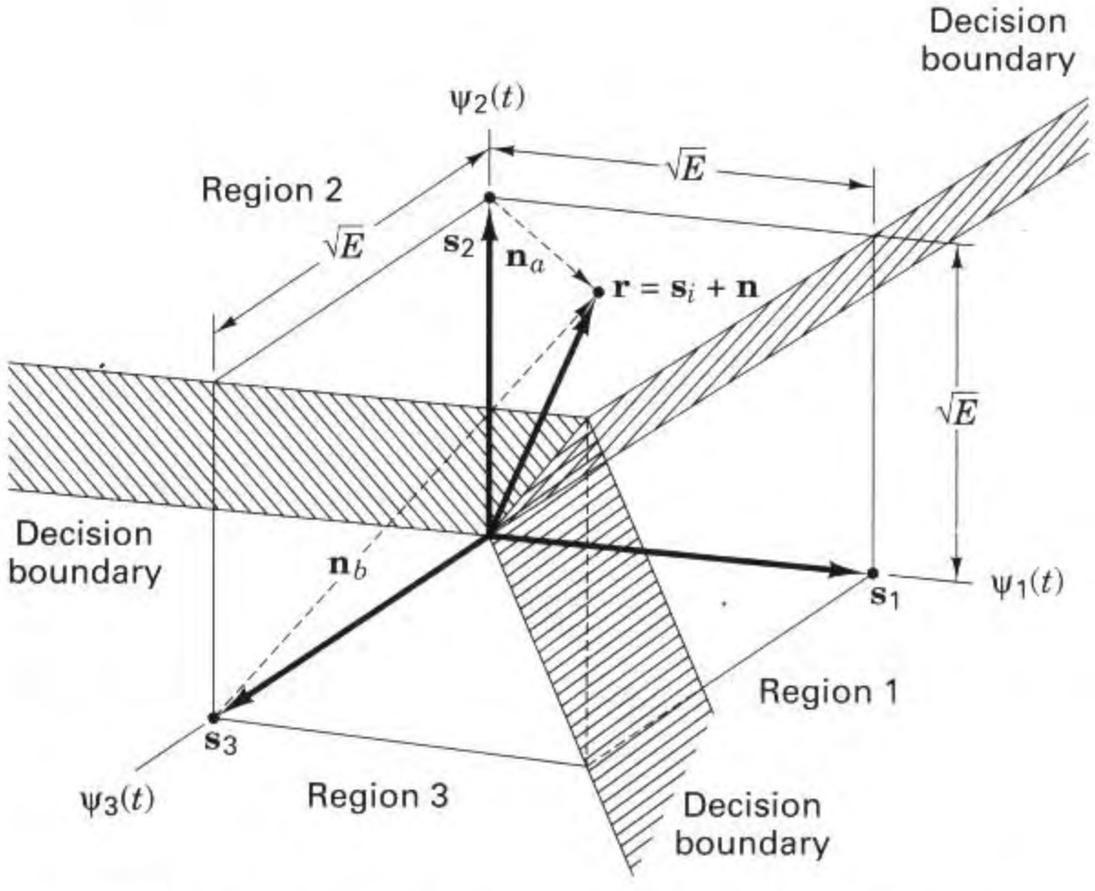


Figure 4.14 Partitioning the signal space for a 3-ary FSK signal.

found. In Figure 4.14, a received signal vector $\mathbf{r}$ is shown in region 2. Using the decision rule stated above, the detector classifies $\mathbf{r}$ as signal $\mathbf{s}_2$. Since the noise is a Gaussian random vector, there is a probability greater than zero that $\mathbf{r}$ could have been produced by some signal other than $\mathbf{s}_2$. For example, if the transmitter had sent $\mathbf{s}_2$, then $\mathbf{r}$ would be the sum of signal plus noise, $\mathbf{s}_2 + \mathbf{n}_a$, and the decision to choose $\mathbf{s}_2$ is correct; however, if the transmitter had actually sent $\mathbf{s}_3$, then $\mathbf{r}$ would be the sum of signal plus noise, $\mathbf{s}_3 + \mathbf{n}_b$, and the decision to select $\mathbf{s}_2$ is an error. The error performance of coherently detected FSK systems is treated in Section 4.7.3.

Example 4.2 Received Phase as a Function of Propagation Delay

(a) In Figure 4.8, the schematics fail to indicate where the correlator reference signals come from. One might think that they are known for all time and stored in memory until needed. Under some controlled circumstances, it is conceivable that the receiver might predict some expected value of the arriving signal's amplitude or its frequency within tolerable limits. But the one parameter that cannot be estimated without special help is the received signal phase. The most popular way to achieve phase estimation is through the use of circuitry called a *phase-locked loop* (PLL). This carrier-recovery method locks on to the arriving carrier wave (or recreates it) and estimates its phase. To show how futile it would be to predict the phase without a PLL, consider the mobile radio link shown in Figure 4.15. In the figure, a mobile user is positioned at point A, at a distance d from the base station, such that the propagation delay is T_d. Using complex notation, the transmitted waveform emanating from the transmitter can be described as $s(t) = \exp(j2\pi f_0 t)$. Let us take f_0 equal to 1 GHz. Neglecting noise, the waveform received at the base station can be denoted as $r(t) = \exp[j2\pi f_0(t + T_d)]$. If the mobile user moves in-line away from the base station to point B, or in-line toward the base station to point C, what is the minimum distance of movement that will cause a 2π phase rotation of the received waveform?

(b) Do we really care about a 2π phase rotation? Of course not, because the received phasor would then be located at exactly the point at which it was initially postulated with the user located at point A. But let us ask, What is the minimum distance that will cause a $\pi/2$ phase rotation (say a lag of $\pi/2$)? Had the receiver predicted that the received phasor would correspond to $r(t)$, as denoted in part (a),

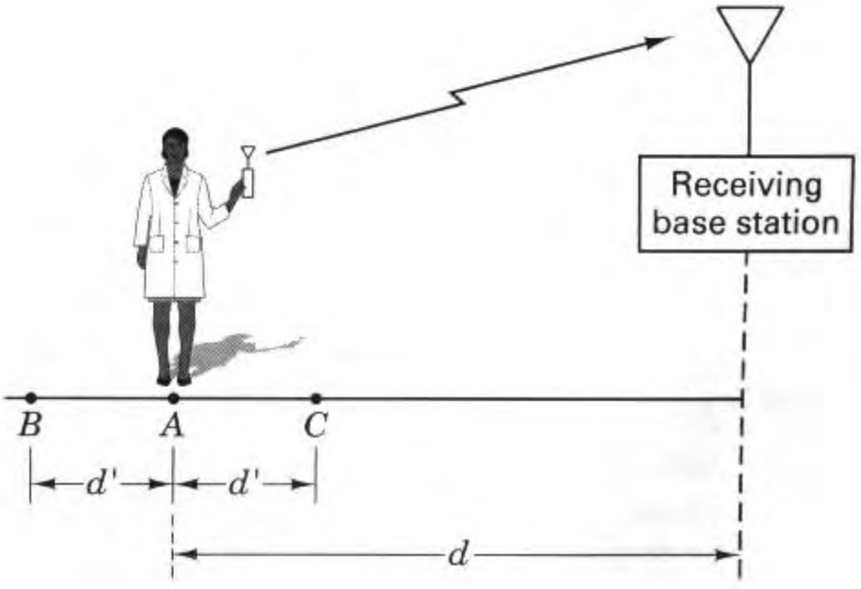


Figure 4.15 Mobile radio link.

but instead the lag caused the received phasor to appear as exp $[j2\pi f_0(t + T_d) - \pi/2]$, then the detecting correlator would yield a zero output. This is so because

$$\int_0^T \cos \omega_0 t \cos\left(\omega_0 t - \frac{\pi}{2}\right) dt = \int_0^T \cos \omega_0 t \sin \omega_0 t \, dt = 0$$

Find the user's minimum distance movement that will cause a $\pi/2$ phase rotation.

Solution

(a) Initially, let $t = 0$, so that when the mobile user is located at point A, the received phasor at the base station can be expressed as $r(t) = \exp(j2\pi f_0 T_d)$. Then, after the user's movement to point B, the received (further delayed) phasor $r_d(t = T_d + T_d')$, can be written as $r_d(t) = \exp[j2\pi f_0(T_d + T_d')]$. The minimum delay time $T_{d'}$ corresponding to a 2π (one wavelength) phasor rotation is $T_d' = 1/f_0 = 10^{-9}$ second. Therefore, the minimum distance for such a rotation (assuming ideal electromagnetic propagation at the speed of light) is

$$d' = \frac{c}{f_0} = 10^8 \text{ m/s} \times 10^{-9} \text{ s} = 0.3 \text{ m}$$

(b) Thus, for a $\pi/2$ phasor rotation, the minimum distance is

$$d'' = \frac{d'}{4} = \frac{0.3 \text{ m}}{4} = 7.5 \text{ cm}$$

It should be clear that even if a transmitter and receiver are located on fixed towers, a small amount of wind movement can bring about complete uncertainty regarding phase. If we scale our example from a frequency of 1 GHz to that of 10 GHz, the minimum distance scales from 7.5 cm to 0.75 cm. Very often we might want to avoid building receivers with PLLs for carrier recovery. The results of this example might then motivate us to ask, How will the error performance suffer if phase information is not used in the detection process? In other words, how will the system fare if the detection is performed noncoherently? We address this question in the sections that follow.

4.5 NONCOHERENT DETECTION

4.5.1 Detection of Differential PSK

The name *differential* PSK (DPSK) sometimes needs clarification because two separate aspects of the modulation/demodulation format are being referred to: the encoding procedure and the detection procedure. The term *differential encoding* refers to the procedure of encoding the data differentially; that is, the presence of a binary one or zero is manifested by the symbol's similarity or difference when compared with the preceding symbol. The term *differentially coherent detection* of differentially encoded PSK, the usual meaning of DPSK, refers to a detection scheme often classified as noncoherent because it does not require a reference in phase with the received carrier. Occasionally, differentially encoded PSK is *coherently* detected. This will be discussed in Section 4.7.2.

With noncoherent systems, no attempt is made to determine the actual value of the phase of the incoming signal. Therefore, if the transmitted waveform is

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos [\omega_0 t + \theta_i(t)] \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix}$$

the received signal can be characterized by

$$r(t) = \sqrt{\frac{2E}{T}} \cos [\omega_0 t + \theta_i(t) + \alpha] + n(t) \qquad \begin{matrix} 0 \le t \le T \\ i = 1, \ldots, M \end{matrix} \tag{4.41}$$

where α is an arbitrary constant and is typically assumed to be a random variable uniformly distributed between zero and 2π, and $n(t)$ is an AWGN process.

For coherent detection, matched filters (or their equivalents) are used; for noncoherent detection, this is not possible because the matched filter output is a function of the unknown angle α. However, if we assume that α varies slowly relative to two period times ($2T$), the phase difference between two successive waveforms $\theta_j(T_1)$ and $\theta_k(T_2)$ is independent of α; that is,

$$[\theta_k(T_2) + \alpha] - [\theta_j(T_1) + \alpha] = \theta_k(T_2) - \theta_j(T_1) = \phi_i(T_2) \tag{4.42}$$

The basis for *differentially coherent detection* of differentially encoded PSK (DPSK) is as follows. The carrier phase of the previous signaling interval can be used as a phase reference for demodulation. Its use requires *differential encoding* of the message sequence at the transmitter since the information is carried by the difference in phase between two successive waveforms. To send the ith message ($i = 1, 2, \ldots, M$), the present signal waveform must have its phase advanced by $\phi_i = 2\pi i/M$ radians over the previous waveform. The detector, in general, calculates the coordinates of the incoming signal by correlating it with locally generated waveforms, such as $\sqrt{2/T} \cos \omega_0 t$ and $\sqrt{2/T} \sin \omega_0 t$. the detector then measures the angle between the currently received signal vector and the previously received signal vector, as illustrated in Figure 4.16.

In general, DPSK signaling performs less efficiently than PSK, because the errors in DPSK tend to propagate (to adjacent symbol times) due to the correlation between signaling waveforms. One way of viewing the difference between PSK and

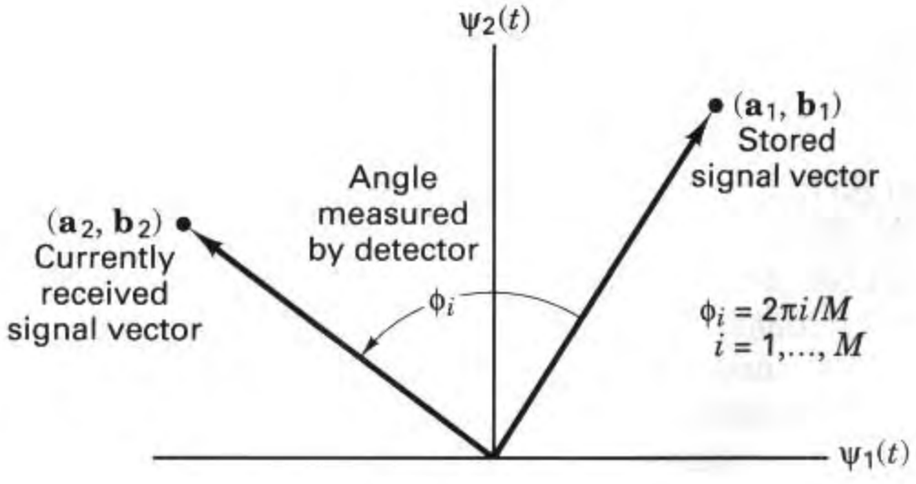


Figure 4.16 Signal space for DPSK.

DPSK is that the former compares the received signal with a clean reference; in the latter, however, two noisy signals are compared with each other. We might say that there is twice as much noise associated with DPSK signaling compared to PSK signaling. Consequently, as a first guess, we might estimate that the error probability for DPSK is approximately two times (3 dB) worse than PSK; this degradation decreases rapidly with increasing signal-to-noise ratio. The trade-off for this performance loss is reduced system complexity. The error performance for the detection of DPSK is treated in Section 4.7.5.

4.5.2 Binary Differential PSK Example

The essence of differentially coherent detection in DPSK is that the identity of the data is inferred from the changes in phase from symbol to symbol. Therefore, because the data are detected by differentially examining the waveform, the transmitted waveform must first be encoded in a differential fashion. Figure 4.17a illustrates a differential encoding of a binary message data stream $m(k)$, where k is the sample time index. The differential encoding starts (third row in the figure) with the first bit of the code-bit sequence $c(k = 0)$, chosen arbitrarily (here taken to be a one). Then the sequence of encoded bits $c(k)$ can, in general, be encoded in one of two ways:

$$c(k) = c(k - 1) \oplus m(k) \tag{4.43}$$

or

$$c(k) = \overline{c(k - 1) \oplus m(k)} \tag{4.44}$$

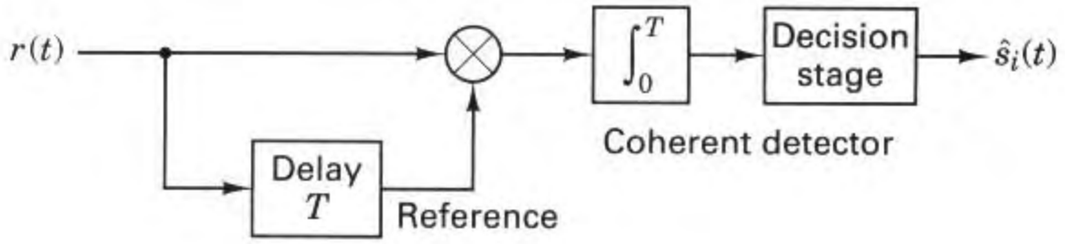
where the symbol $\oplus$ represents modulo-2 addition (defined in Section 2.9.3) and the overbar denotes complement. In Figure 4.17a the differentially encoded message was obtained by using Equation (4.44). In other words, the present code bit $c(k)$ is a one if the message bit $m(k)$ and the prior coded bit $c(k - 1)$ are the same, otherwise, $c(k)$ is a zero. The fourth row translates the coded bit sequence $c(k)$ into the phase shift sequence $\theta(k)$, where a one is characterized by a 180° phase shift, and a zero is characterized by a 0° phase shift.

Figure 4.17b illustrates the binary DPSK detection scheme in block diagram form. Notice that the basic product integrator of Figure 4.7 is the essence of the demodulator; as with coherent PSK, we are still attempting to correlate a received signal with a reference. The interesting difference here is that the reference signal is simply a delayed version of the received signal. In other words, during each symbol time, we are matching a received symbol with the prior symbol and looking for a correlation or an anticorrelation (180° out of phase).

Consider the received signal with phase shift sequence $\theta(k)$ entering the correlator of Figure 4.17b, in the absence of noise. The phase $\theta(k = 1)$ is matched with $\theta(k = 0)$; they have the same value, π; hence the first bit of the detected output is $\hat{m}(k = 1) = 1$. Then $\theta(k = 2)$ is matched with $\theta(k = 1)$; again they have the same value, and $\hat{m}(k = 2) = 1$. Then $\theta(k = 3)$ is matched with $\theta(k = 2)$; they are different, so that $\hat{m}(k = 3) = 0$, and so on.

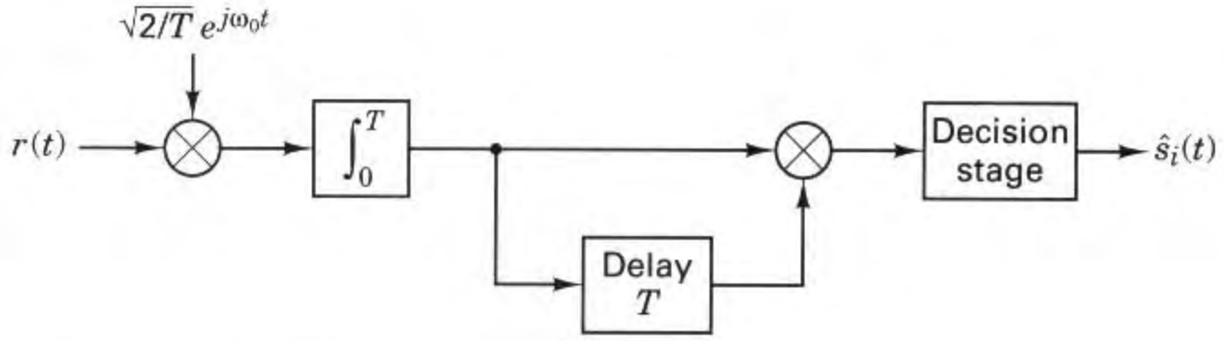
Sample index, k	0	1	2	3	4	5	6	7	8	9	10
Information message, $m(k)$		1	1	0	1	0	1	1	0	0	1
Differentially encoded message (first bit arbitrary), $c(k)$	1	1	1	0	0	1	1	1	0	1	1
Corresponding phase shift, $\theta(k)$	π	π	π	0	0	π	π	π	0	π	π

(a)



Detected message, $\hat{m}(k)$ 1 1 0 1 0 1 1 0 0 1

(b)



(c)

Figure 4.17 Differential PSK (DPSK). (a) Differential encoding. (b) Differentially coherent detection. (c) Optimum differentially coherent detection.

It must be pointed out that the detector in Figure 4.17b is suboptimum [3] in the sense of error performance. The optimum differential detector for DPSK requires a reference carrier in frequency but not necessarily in phase with the received carrier. Hence the optimum differential detector is shown in Figure 4.17c [4]. Its performance is treated in Section 4.7.5. In Figure 4.17c, the reference signal is shown in complex form ($\sqrt{2/T}\, e^{j\omega_0 t}$) to indicate that a quadrature implementation using both I and Q references are required (see Section 4.6.1).

4.5.3 Noncoherent Detection of FSK

A detector for the *noncoherent* detection of FSK waveforms described by Equation (4.8) can be implemented with correlators similar to those shown in Figure 4.7. However, the hardware must be configured as an *energy detector,* without exploiting phase measurements. For this reason, the noncoherent detector typically requires twice as many channel branches as the coherent detector. Figure 4.18 illustrates the in-phase (I) and quadrature (Q) channels used to detect a binary FSK (BFSK) signal set noncoherently. Notice that the upper two branches are configured to detect the signal with frequency ω_1; the reference signals are $\sqrt{2/T}\cos\omega_1 t$ for the I branch and $\sqrt{2/T}\sin\omega_1 t$ for the Q branch. Similarly, the lower two branches are configured to detect the signal with frequency ω_2; the reference signals are $\sqrt{2/T}\cos\omega_2 t$ for the I branch and $\sqrt{2/T}\sin\omega_2 t$ for the Q branch. Imagine that the received signal $r(t)$, by chance alone, is exactly of the form $\cos\omega_1 t + n(t)$; that is, the phase is exactly zero, and thus the signal component of the received signal exactly matches the top-branch reference signal with regard to frequency and phase. In that event, the product integrator of the top branch should yield the maximum output. The second branch should yield a near-zero output (integrated zero-mean noise), since its reference signal $\sqrt{2/T}\sin\omega_1 t$ is orthogonal to the sig-

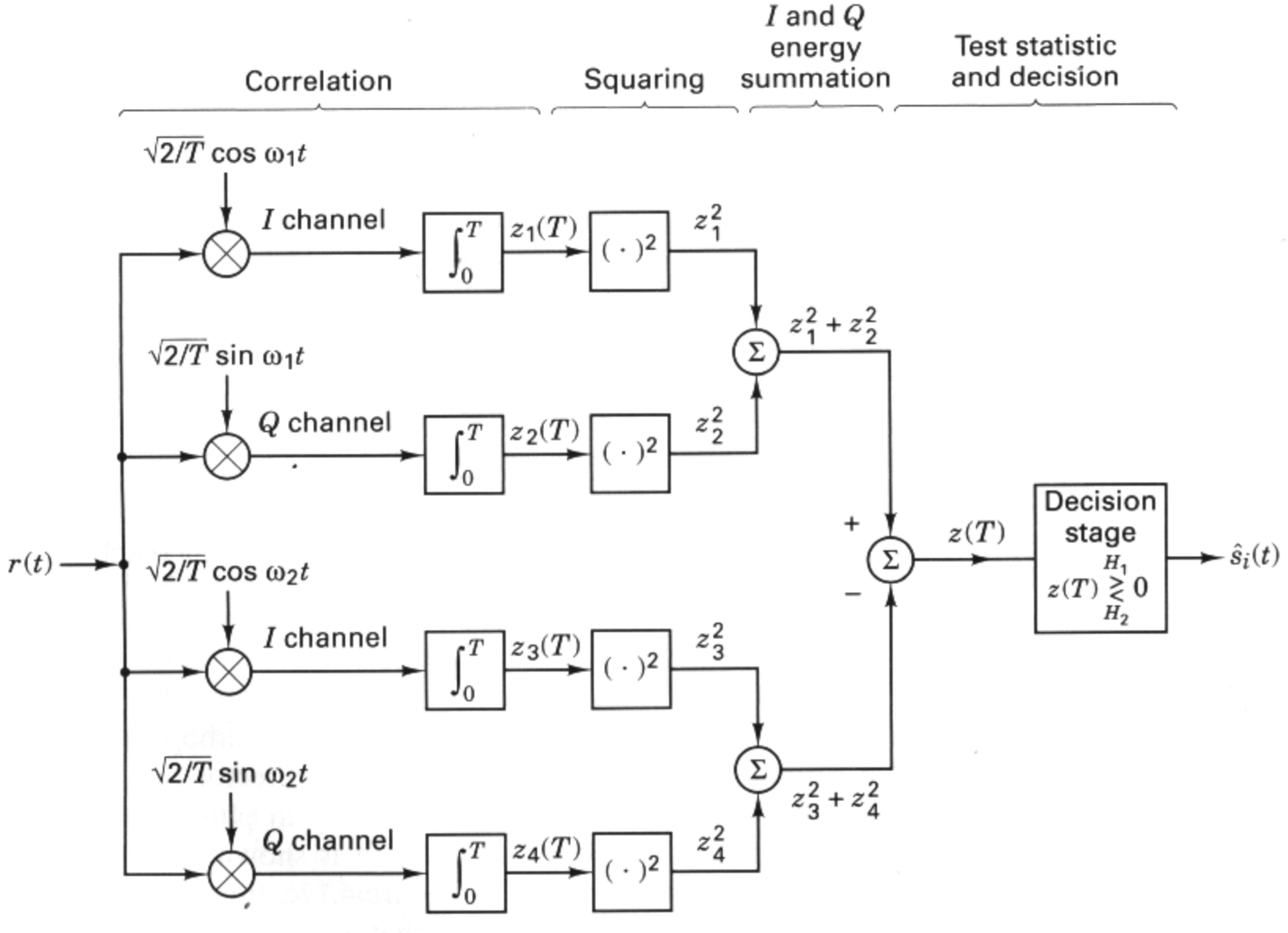


Figure 4.18 Quadrature receiver.

nal component of $r(t)$. For orthogonal signaling (see Section 4.5.4), the third and fourth branches should also yield near-zero outputs, since their ω_2 reference signals are also orthogonal to the signal component of $r(t)$.

Now imagine a different scenario: Suppose that by chance alone, the received signal $r(t)$ is of the form $\sin \omega_1 t + n(t)$. In that event, the second branch in Figure 4.18 should yield the maximum output, while the others should yield near-zero outputs. In actual practice, the most likely scenario is that $r(t)$ is of the form $\cos(\omega_1 t + \phi) + n(t)$; that is, the incoming signal will *partially* correlate with the $\cos \omega_1 t$ reference and *partially* correlate with the $\sin \omega_1 t$ reference. Hence a noncoherent quadrature receiver for orthogonal signals requires an I and Q branch for each candidate signal in the signaling set. In Figure 4.18, the blocks following the product integrators perform a squaring operation to prevent the appearance of any negative values. Then for each signal class in the set (two in this binary example), z_1^2 from the I channel and z_2^2 from the Q channel are added. The final stage forms the test statistic $z(T)$ and chooses the signal with frequency ω_1 or the signal with frequency ω_2, depending upon which pair of energy detectors yields the maximum output.

Another possible implementation for noncoherent FSK detection uses bandpass filters—centered at $f_i = \omega_i/2\pi$ with bandwidth $W_f = 1T$—followed by *envelope detectors,* as shown in Figure 4.19. An envelope detector consists of a rectifier and a low-pass filter. The detectors are matched to the *signal envelopes* and not to the signals themselves. The phase of the carrier is of no importance in defining the envelope; hence no phase information is used. In the case of binary FSK, the decision as to whether a one or a zero was transmitted is made on the basis of which of two envelope detectors has the largest amplitude at the moment of measurement. Similarly, for a multiple frequency shift-keying (MFSK) system, the decision as to which of M signals was transmitted is made on the basis of which of the M envelope detectors has the maximum output.

Even though the envelope detector block diagram of Figure 4.19 looks functionally more simple than the quadrature receiver shown in Figure 4.18, the use of

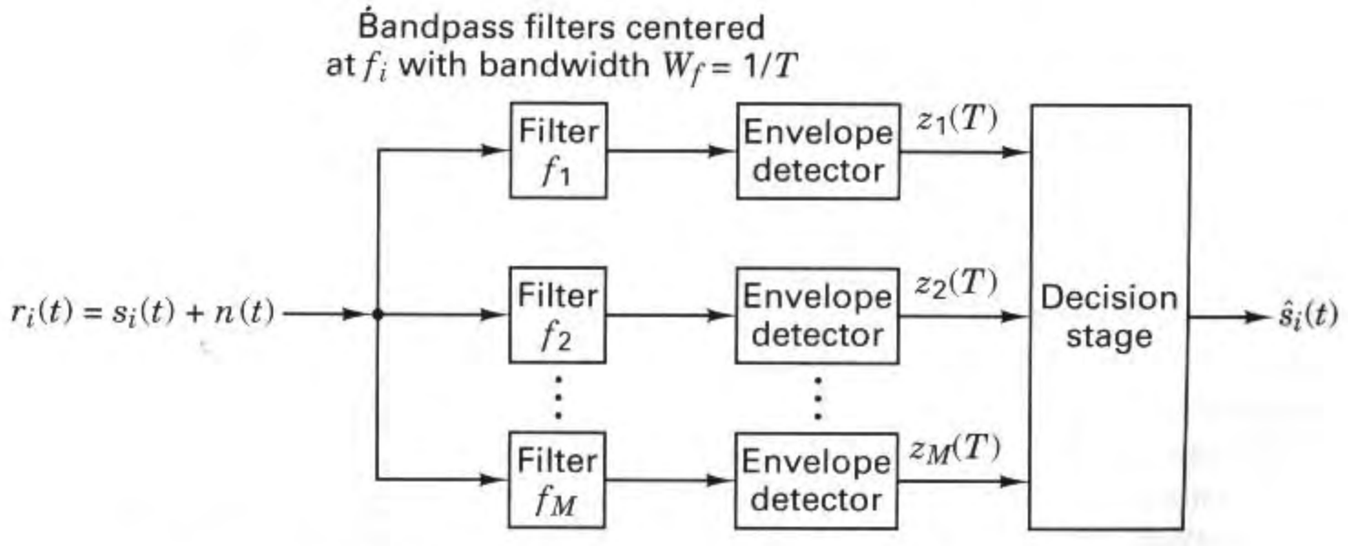


Figure 4.19 Noncoherent detection of FSK using envelope detectors.

(analog) filters usually results in the envelope detector design having greater weight and cost than the quadrature receiver. Quadrature receivers can be implemented digitally; thus, with the advent of large-scale integrated (LSI) circuits, they are often the preferred choice for noncoherent detectors. The detector shown in Figure 4.19 can also be implemented digitally by performing discrete Fourier transformations instead of using analog filters, but such a design is usually more complex than a digital implementation of the quadrature receiver.

4.5.4 Required Tone Spacing for Noncoherent Orthogonal FSK Signaling

Frequency shift keying (FSK) is usually implemented as orthogonal signaling, but not all FSK signaling is orthogonal. How can we tell if the tones in a signaling set form an orthogonal set? For example, suppose we have two tones $f_1 = 10{,}000$ Hz and $f_2 = 11{,}000$ Hz. Are they orthogonal to each other? In other words, do they fulfill the criterion (as set forth in Equation 3.69) that they are uncorrelated over a symbol time T? We do not have enough information to answer that question yet. Tones f_1 and f_2 manifest orthogonality if, for a transmitted tone at f_1, the sampled envelope of the receiver output filter tuned to f_2 is zero (i.e., no crosstalk). A property that insures such orthogonality between tones in an FSK signaling set states that any pair of tones in the set must have a frequency separation that is a multiple of $1/T$ hertz. (This will be proven below, in Example 4.3.) A tone with frequency f_i that is switched on for a symbol duration of T seconds and then switched off, such as the FSK tone described in Equation (4.8), can be analytically described by

$$s_i(t) = (\cos 2\pi f_i t) \text{ rect } (t/T)$$

$$\text{where rect } (t/T) = \begin{cases} 1 & \text{for } -T/2 \leq t \leq T/2 \\ 0 & \text{for } |t| > T/2 \end{cases}$$

From Table A.1, the Fourier transform of $s_i(t)$ is

$$\mathcal{F}\{s_i(t)\} = T \text{ sinc } (f - f_i)T$$

where the sinc function is as defined in Equation (1.39). The spectra of two such adjacent tones—tone 1 with frequency f_1 and tone 2 with frequency f_2—are plotted in Figure 4.20.

4.5.4.1 Minimum Tone Spacing and Bandwidth

In order for a noncoherently detected tone to manifest a maximum output signal at its associated receiver filter and a zero-output signal at any neighboring filter (as implemented in Figure 4.19), the peak of the spectrum of tone 1 must coincide with one of the zero crossings of the spectrum of tone 2, and similarly, the peak of the spectrum of tone 2 must coincide with one of the zero crossings of the spectrum of tone 1. The frequency difference between the center of the spectral main lobe and the first zero crossing represents the *minimum required spacing.* With noncoherent detection, this corresponds to a minimum tone separation of $1/T$ hertz, as can be seen in Figure 4.20. Even though FSK signaling entails the trans-

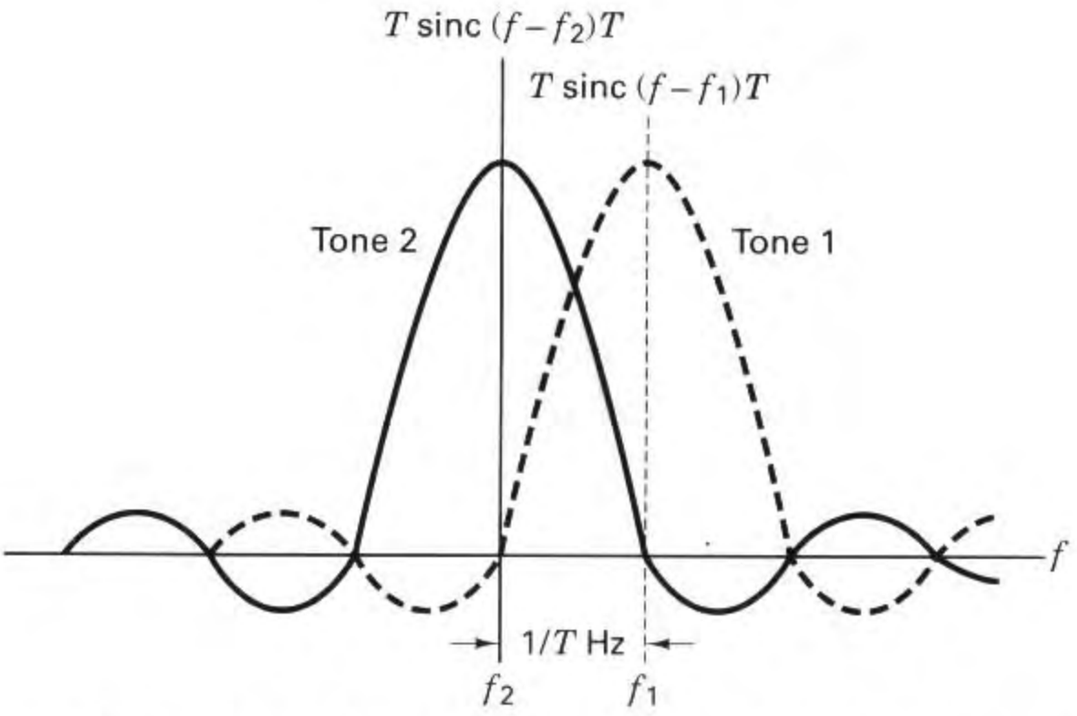


Figure 4.20 Minimum tone spacing for noncoherently detected orthogonal FSK signaling.

mission of just one single-sideband tone during each symbol time, when we speak of the signal bandwidth, we mean the amount of spectrum that needs to be available for the entire range of all the tones in the M-ary set. Thus for FSK, the bandwidth requirements are related to the spectral spacing between the tones. From a group of adjacent tones, the spectrum associated with just one tone can be considered to make up one-half of the tone spacing on each side of the tone. Thus, for the binary FSK case shown in Figure 4.20, the signaling bandwidth is equal to the spectrum that separates the tone centers plus one-half the tone spacing on each side of the set. This amounts to two times the tone spacing. Extrapolating this approach to the M-ary case, the bandwidth of noncoherently detected orthogonal MFSK is equal to M/T.

Thus far, we have only addressed noncoherently detected orthogonal FSK. Is the minimum tone-spacing criterion (and hence the bandwidth) different if the signaling is detected coherently? Indeed, it is. As we will see below, in Example 4.3, the minimum tone spacing is reduced to $1/2T$ when coherent detection is used.

4.5.4.2 Dual Relationships

The engineering concept of duality can be defined as follows: Two processes (functions, elements, or systems) are *dual* to each other if their mathematical relationships are identical, even though they are described with different variables (e.g., time and frequency). Consider the case of FSK signaling, where rectangular-shaped keying corresponds to sinc (fT) shaped tones, as seen in Figure 4.20. A given tone duration gives rise to the minimum frequency spacing between tones that would be needed to achieve orthogonality. This frequency-domain relationship has as its dual counterpart a time-domain relationship—the case of pusle sig-

naling, where a rectangular-shaped bandwidth corresponds to sinc (t/T) shaped pulses, as seen in Figure 3.16b. A given bandwidth gives rise to the minimum time spacing between pulses that would be needed to achieve zero ISI.

Example 4.3 Minimum Tone Spacing for Orthogonal FSK

Consider two waveforms $\cos(2\pi f_1 t + \phi)$ and $\cos 2\pi f_2 t$ to be used for *noncoherent* FSK-signaling, where $f_1 > f_2$. The symbol rate is equal to $1/T$ symbols/s, where T is the symbol duration and ϕ is a constant arbitrary angle from 0 to 2π.

(a) Prove that the minimum tone spacing for *noncoherently detected* orthogonal FSK-signaling is $1/T$.
(b) What is the minimum tone spacing for *coherently detected* orthogonal FSK signaling?

Solution

(a) For the two waveforms to be orthogonal, they must fulfill the orthogonality constraint of Equation (3.69):

$$\int_0^T \cos(2\pi f_1 t + \phi) \cos 2\pi f_2 t \, dt = 0 \tag{4.45}$$

Using the basic trigonometric identities shown in Equations (D.6) and (D.1) to (D.3), we can write Equation (4.45) as

$$\cos\phi \int_0^T \cos 2\pi f_1 t \cos 2\pi f_2 t \, dt - \sin\phi \int_0^T \sin 2\pi f_1 t \cos 2\pi f_2 t \, dt = 0 \tag{4.46}$$

so that

$$\cos\phi \int_0^T [\cos 2\pi(f_1 + f_2)t + \cos 2\pi(f_1 - f_2)t] \, dt - \sin\phi \int_0^T [\sin 2\pi(f_1 + f_2)t + \sin 2\pi(f_1 - f_2)t] \, dt = 0 \tag{4.47}$$

which, in turn, yields

$$\cos\phi \left[\frac{\sin 2\pi(f_1 + f_2)t}{2\pi(f_1 + f_2)} + \frac{\sin 2\pi(f_1 - f_2)t}{2\pi(f_1 - f_2)} \right]_0^T + \sin\phi \left[\frac{\cos 2\pi(f_1 + f_2)t}{2\pi(f_1 + f_2)} + \frac{\cos 2\pi(f_1 - f_2)t}{2\pi(f_1 - f_2)} \right]_0^T = 0 \tag{4.48}$$

or

$$\cos\phi \left[\frac{\sin 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} + \frac{\sin 2\pi(f_1 - f_2)T}{2\pi(f_1 - f_2)} \right] + \sin\phi \left[\frac{\cos 2\pi(f_1 + f_2)T - 1}{2\pi(f_1 + f_2)} + \frac{\cos 2\pi(f_1 - f_2)T - 1}{2\pi(f_1 - f_2)} \right] = 0 \tag{4.49}$$

We can assume that $f_1 + f_2 \gg 1$ and can thus make the following approximation:

$$\frac{\sin 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} \approx \frac{\cos 2\pi(f_1 + f_2)T}{2\pi(f_1 + f_2)} \approx 0 \tag{4.50}$$

Then, combining Equations (4.49) and (4.50), we can write

$$\cos\phi \sin 2\pi(f_1 - f_2)T + \sin\phi\,[\cos 2\pi(f_1 - f_2)T - 1] \approx 0 \tag{4.51}$$

Note that for arbitrary ϕ, the terms in Equation (4.51) can sum to zero only when $\sin 2\pi(f_1 - f_2)T = 0$, and simultaneously $\cos 2\pi(f_1 - f_2)T = 1$.

Since

$$\sin x = 0 \qquad \text{for } x = n\pi$$

and

$$\cos x = 1 \qquad \text{for } x = 2k\pi$$

where n and k are integers, then both $\sin x = 0$ and $\cos x = 1$ occur simultaneously when $n = 2k$. From Equation (4.51), for arbitrary ϕ, we can therefore write

$$2\pi(f_1 - f_2)T = 2k\pi$$

or (4.52)

$$f_1 - f_2 = \frac{k}{T}$$

Thus the minimum tone spacing for *noncoherent* FSK signaling occurs for $k = 1$, in which case we write

$$f_1 - f_2 = \frac{1}{T} \tag{4.53}$$

Recall the question posed earlier. Having two tones f_1 = 10,000 hertz and f_2 = 11,000 hertz, we asked, Are they orthogonal? Now we have sufficient information to answer that question. The answer depends on the speed of the FSK signaling. If the tones are being keyed (switched) at the rate of 1,000 symbols/s and noncoherently detected, then they are orthogonal. If they are being keyed faster—say, at the rate of 10,000 symbols/s—they are not orthogonal.

(b) For the noncoherent detection in part (a) of this example, the tone spacings that rendered the signals orthogonal was found by satisfying Equation (4.45) for any arbitrary phase. In this case, however, for coherent detection, the tone spacings needed for orthogonality are found by setting $\phi = 0$. This is because we know the phase of the received signal (from our phase-locked loop estimate). This received signal will be correlated with each reference signal using the same phase estimate for the reference signals. We can now rewrite Equation (4.51) with $\phi = 0$, which gives

$$\sin 2\pi(f_1 - f_2)T = 0 \tag{4.54}$$

or

$$f_1 - f_2 = \frac{n}{2T} \tag{4.55}$$

Thus the minimum tone spacing for *coherent* FSK signaling occurs for $n = 1$ as follows:

$$f_1 - f_2 = \frac{1}{2T} \tag{4.56}$$

Therefore, for the same symbol rate, coherently detected FSK can occupy less bandwidth than noncoherently detected FSK and still retain orthogonal signaling. We can say that coherent FSK is more *bandwidth efficient.* (The subject of bandwidth efficiency is addressed in greater detail in Chapter 9.)

The required tone spacings are now closer than in part (a) because when we align two periodic waveforms so that their starting phases are the same, we achieve orthogonality by virtue of an even-versus-odd symmetry in the respective waveforms over one symbol time. This is unlike the way orthogonality was achieved in part (a), where we paid no attention to phase. In the coherent case here, the phase alignment in the correlator stages means that we can bring the tones closer together in frequency and still maintain orthogonality among the set of FSK tones. Prove it to yourself by plotting two sine waves (or cosine waves or square waves). Start them off at the same phase (0 radians is convenient). Using quadrille paper, choose a simple time scale to represent one symbol time T. Then, plot a tone at one cycle per T. Below that, with the same starting phase, plot another tone at one-and-a-half cycles per T. Perform product-summation over period T, and verify that these waveforms (spaced $1/2T$ hertz apart) are indeed orthogonal.

4.6 COMPLEX ENVELOPE

The description of real-world modulators and demodulators is facilitated by the use of complex notation which began in Section 4.2.1 and continues here. Any real bandpass waveform $s(t)$ can be represented using complex notation as

$$s(t) = \text{Re}\{g(t)e^{j\omega_0 t}\} \tag{4.57}$$

where $g(t)$ is known as the *complex envelope,* expressed as

$$g(t) = x(t) + jy(t) = |g(t)|e^{j\theta(t)} = R(t)e^{j\theta(t)} \tag{4.58}$$

The magnitude of the complex envelope is then

$$R(t) = |g(t)| = \sqrt{x^2(t) + y^2(t)} \tag{4.59}$$

and its phase is

$$\theta(t) = \tan^{-1}\frac{y(t)}{x(t)} \tag{4.60}$$

With respect to Equation (4.57), we can call $g(t)$ the baseband message or data in complex form, and $e^{j\omega_0 t}$ the carrier wave in complex form. The product of these two represents modulation, and $s(t)$, the real part of this product, is the transmitted waveform. Therefore, using Equations (4.4), (4.57), and (4.58), we can express $s(t)$ as follows:

$$
\begin{aligned}
s(t) &= \mathrm{Re}\{[x(t) + jy(t)][\cos \omega_0 t + j \sin \omega_0 t]\} \\
&= x(t) \cos \omega_0 t - y(t) \sin \omega_0 t
\end{aligned} \tag{4.61}
$$

Note that the modulation of signals expressed in the general form of $(a + jb)$ times $(c + jd)$ yields a waveform with a sign inversion (at the quadrature term of the carrier wave) of the form *ac-bd.*

4.6.1 Quadrature Implementation of a Modulator

Consider an example of a baseband waveform $g(t)$, described by a sequence of ideal pulses $x(t)$ and $y(t)$ appearing at discrete times $k = 1, 2, \ldots$. Thus, $g(t)$, $x(t)$, and $y(t)$ in Equation (4.58) can be written as g_k, x_k, and y_k, respectively. Let the pulse-amplitude values be $x_k = y_k = 0.707A$. The complex envelope can then be expressed in discrete form as

$$
g_k = x_k + jy_k = 0.707A + j0.707A \tag{4.62}
$$

We know from complex algebra that $j = \sqrt{-1}$, but from a practical point of view, the j term can be regarded as a "flag," reminding us that we may not use ordinary addition in combing the terms in Equation (4.62). In preparation for the next step, inphase and quadrature modulation, x_k and y_k are treated as an ordered pair. A quadrature-type modulator is shown in Figure 4.21, where we see that the x_k pulse is multiplied by $\cos \omega_0 t$ (inphase component of the carrier wave), and the y_k pulse is multiplied by $\sin \omega_0 t$ (quadrature component of the carrier wave). The modulation process can be described succinctly as multiplying the complex envelope by $e^{j\omega_0 t}$, and then transmitting the real part of the product. Hence, we write

$$
\begin{aligned}
s(t) &= \mathrm{Re}\{g_k e^{j\omega_0 t}\} \\
&= \mathrm{Re}\{(x_k + jy_k)(\cos \omega_0 t + j \sin \omega_0 t)\} \\
&= x_k \cos \omega_0 t - y_k \sin \omega_0 t \\
&= 0.707A \cos \omega_0 t - 0.707A \sin \omega_0 t \\
&= A \cos\left(\omega_0 t + \frac{\pi}{4}\right)
\end{aligned} \tag{4.63}
$$

Again, note that the quadrature term of the carrier wave has undergone a sign inversion in the modulation process. If we use $0.707A \cos \omega_0 t$ as a reference, then the transmitted waveform $s(t)$ in Equation (4.63) leads the reference waveform by $\pi/4$.

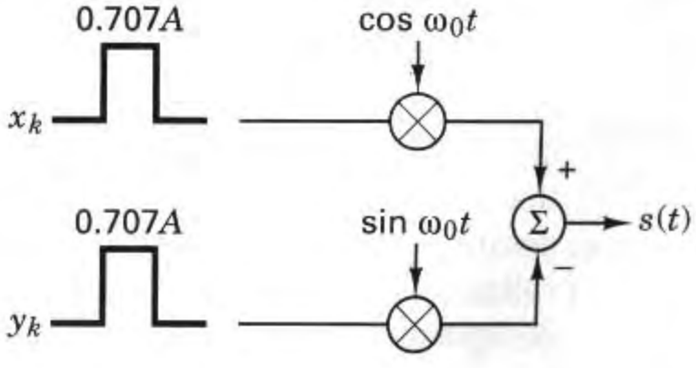


Figure 4.21 Quadrature type modulator.

Similarly, if we consider $-0.707A \sin \omega_0 t$ as the reference, then $s(t)$ in Equation (4.63) lags the reference waveform by $\pi/4$. These relationships are illustrated in Figure 4.22.

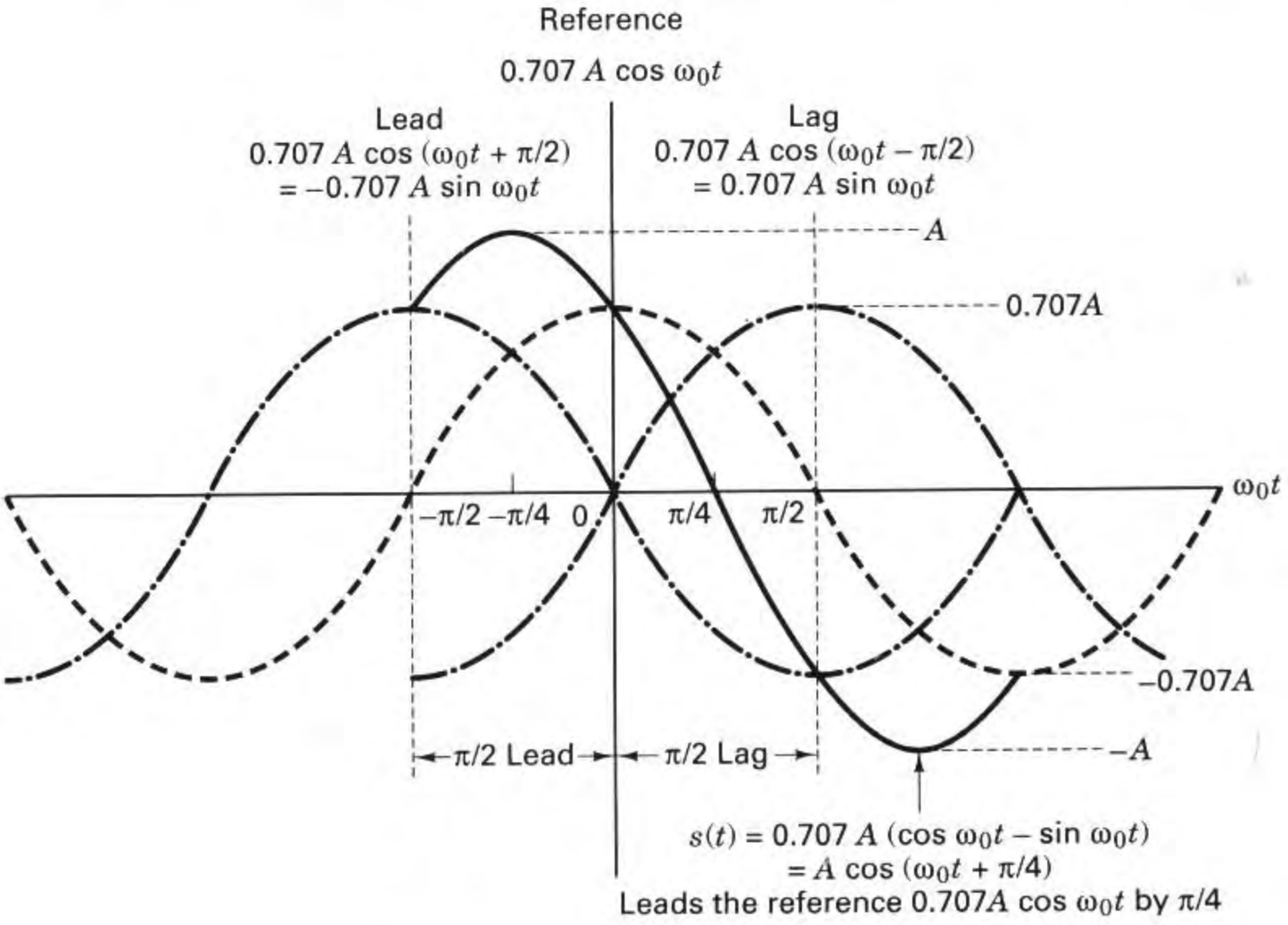


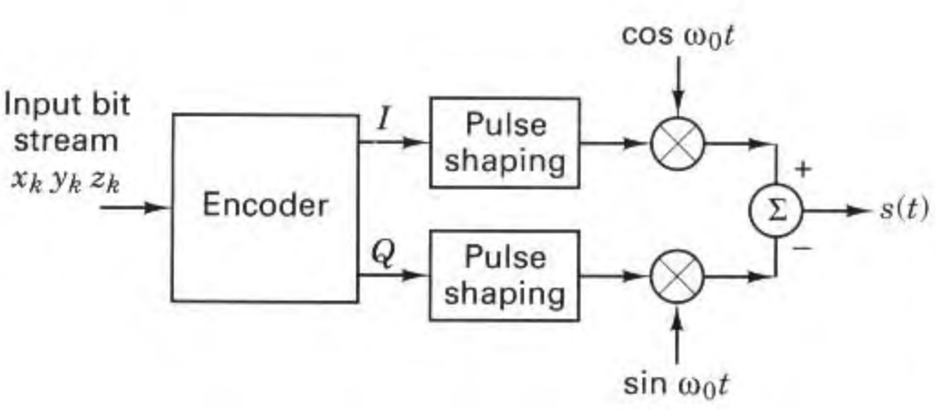
Figure 4.22 Lead/Lag relationships of sinusoids.

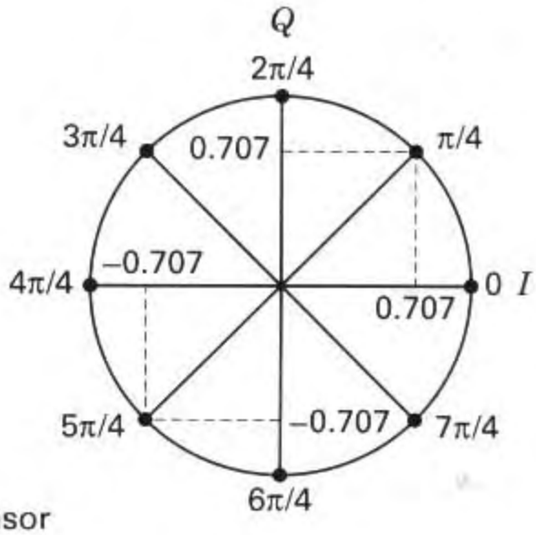
4.6.2 D8PSK Modulator Example

Figure 4.23 depicts a quadrature implementation of a differential 8-PSK (D8PSK) modulator. Because the modulation is 8-ary, we assign a 3-bit message (x_k, y_k, z_k) to each phase $\Delta\phi_k$. Because the modulation is differential, at each kth transmission time we send a data phasor ϕ_k, which can be expressed as

$$\phi_k = \Delta\phi_k + \phi_{k-1} \tag{4.64}$$

The process of adding the current message-to-phase assignment $\Delta\phi_k$ to the prior data phase ϕ_{k-1} provides for the differential encoding of the message. A sequence of phasors created by following Equation (4.64) yields a similar differential encoding as the procedures described in Section 4.5.2. You might notice in Figure 4.23 that the assignment of 3-bit message sequences to $\Delta\phi_k$ does not proceed along the natural binary progression from 000 to 111. There is a special code being used here called a *Gray code.* (The benefits that such a binary to M-ary assignment provides is explained in Section 4.9.4.)





Data encoding

x_k	y_k	z_k	$\Delta\phi_k$
0	0	0	0
0	0	1	$\pi/4$
0	1	1	$2\pi/4$
0	1	0	$3\pi/4$
1	1	0	$4\pi/4$
1	1	1	$5\pi/4$
1	0	1	$6\pi/4$
1	0	0	$7\pi/4$

Differential data phasor

$\phi_k = \phi_{k-1} + \Delta\phi_k$

Let $\phi_0 = 0$	$k = 1$	$k = 2$	$k = 3$	$k = 4$
$x_k\, y_k\, z_k$	110	001	110	010
$\Delta\phi_k$:	π	$\pi/4$	π	$3\pi/4$
ϕ_k:	π	$5\pi/4$	$\pi/4$	π
I:	−1	−0.707	0.707	−1
Q:	0	−0.707	0.707	0

Figure 4.23 Quadrature implementation of a D8PSK modulator.

For the modulator in Figure 4.23, let the input data sequence at times $k = 1, 2, 3, 4$, be equal to 110, 001, 110, 010, respectively. Next, we use the date-encoding table shown in Figure 4.23 and Equation (4.64), with the starting phase at time $k = 0$ to be $\phi_0 = 0$. At time $k = 1$, the differential data phase corresponding to $x_1\, y_1\, z_1 = 110$ is $\phi_1 = 4\pi/4 = \pi$. Taking the magnitude of the rotating phasor to be unity, the inphase (I) and the quadrature (Q) baseband pulses are −1 and 0, respectively. As indicated in Figure 4.23, these pulses are generally shaped with a filter (such as a root-raised-cosine).

For time $k = 2$, the table in Figure 4.23 shows that the message 001 is assigned $\Delta\phi_2 = \pi/4$. Therefore, following Equation (4.64), the second differential data phase is $\phi_2 = \pi + \pi/4 = 5\pi/4$, and for time $k = 2$, the I and Q baseband pulses are $x_k = -0.707$ and $y_k = -0.707$, respectively. The transmitted waveform follows the form of Equation (4.61), rewritten as

$$\begin{aligned} s(t) &= \mathrm{Re}\{(x_k + jy_k)\,(\cos\omega_0 t + j\sin\omega_0 t)\} \\ &= x_k \cos\omega_0 t - y_k \sin\omega_0 t \end{aligned} \tag{4.65}$$

For a signaling set that can be represented on a phase-amplitude plane, such as MPSK or MQAM, Equation (4.65) provides an interesting observation. That is, quadrature implementation of the transmitter transforms all such signaling types

to simple amplitude modulation. Any phasor on the plane is transmitted by amplitude-modulating its inphase and quadrature projections onto the cosine and sine wave components of the carrier, respectively. For ease of notation, we neglect the pulse shaping; that is, we assume that the data pulses have ideal rectangular shapes. Then, using Equation (4.65), for time $k = 2$, where $x_k = -0.707$ and $y_k = -0.707$, the transmitted $s(t)$ can be written as follows:

$$\begin{aligned} s(t) &= -0.707 \cos \omega_0 t + 0.707 \sin \omega_0 t \\ &= \sin\left(\omega_0 t - \frac{\pi}{4}\right) \end{aligned} \tag{4.66}$$

4.6.3 D8PSK Demodulator Example

In the previous section, the quadrature implementation of a modulator began with multiplying the complex envelope (baseband message) by $e^{j\omega_0 t}$, and transmitting the real part of the product $s(t)$, as described in Equation (4.63). Using a similar quadrature implementation, demodulation consists of reversing the process—that is, multiplying the received bandpass waveform by $e^{-j\omega_0 t}$ in order to recover the baseband waveform. The left side of Figure 4.24 shows the modulator of Figure 4.23 in simplified form, and the waveform $s(t) = \sin(\omega_0 t - \pi/4)$ that was launched at time $k = 2$ for that example. We continue this same example in this section, and also show a quadrature implementation of the demodulator on the right side of Figure 4.24.

Notice the subtle difference between the $-\sin \omega_0 t$ term at the modulator and the $-\sin \omega_0 t$ term at the demodulator. At the modulator, the minus sign stems from taking the real part of the complex waveform (product of the complex envelope and complex carrier wave). At the demodulator, $-\sin \omega_0 t$ stems from multiplying the bandpass waveform by the conjugate $e^{-j\omega_0 t}$ of the modulator carrier wave; demodulation is coherent if phase is recovered. To simplify writing the basic relationships of the process, the noise is neglected. After the inphase multiplication by $\cos \omega_0 t$ in the demodulator, we get the signal at point A:

$$\begin{aligned} A &= (-0.707 \cos \omega_0 t + 0.707 \sin \omega_0 t) \cos \omega_0 t \\ &= -0.707 \cos^2 \omega_0 t + 0.707 \sin \omega_0 t \cos \omega_0 t \end{aligned} \tag{4.67}$$

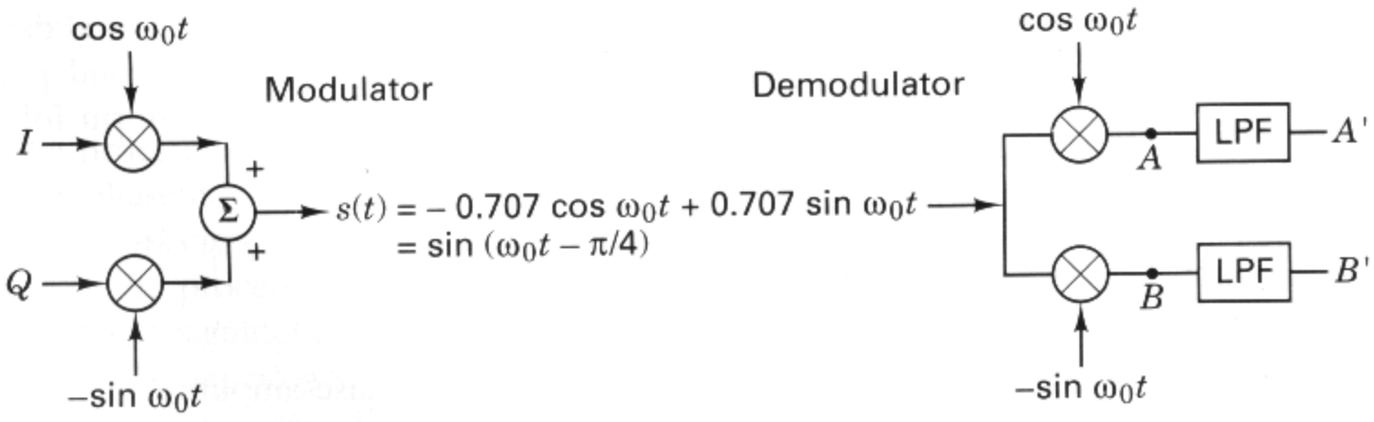


Figure 4.24 Modulator/demodulator example.

Using the trigonometric identities shown in Equation (D.7) and (D.9), we get

$$A = \frac{-0.707}{2}(1 + \cos 2\omega_0 t) + \frac{0.707}{2}\sin 2\omega_0 t \tag{4.68}$$

After filtering with a low-pass filter (LPF) we recover at point A' an ideal negative pulse, as follows:

$$A' = -0.707 \text{ (times a scale factor)} \tag{4.69}$$

Similarly, after the quadrature multiplication by $-\sin \omega_0 t$ in the demodulator, we get the signal at point B, as follows:

$$\begin{aligned} B &= (-0.707 \cos \omega_0 t + 0.707 \sin \omega_0 t)(-\sin \omega_0 t) \\ &= \frac{0.707}{2}\sin 2\omega_0 t - \frac{0.707}{2}(1 - \cos 2\omega_0 t) \end{aligned} \tag{4.70}$$

After the LPF we recover at point B' an ideal negative pulse, as follows:

$$B' = -0.707 \text{ (times a scale factor)} \tag{4.71}$$

Thus, we see from the demodulated and filtered points A' and B' that the differential (ideal) data pulses for the I and Q channels are each equal to -0.707. Since the modulator/demodulator is differential, then for this $k = 2$ example,

$$\Delta\phi_{k=2} = \phi_{k=2} - \phi_{k=1} \tag{4.72}$$

We presume that the demodulator at the earlier time $k = 1$ had properly recovered the signal phase to be π. Then, from Equation (4.72), we can write

$$\Delta\phi_{k=2} = \frac{5\pi}{4} - \pi = \frac{\pi}{4} \tag{4.73}$$

Referring back to the data-encoding table in Figure 4.23, we see that the detected data sequence is $x_2\, y_2\, z_2 = 001$, which corresponds to the data that was sent at time $k = 2$.

4.7 ERROR PERFORMANCE FOR BINARY SYSTEMS

4.7.1 Probability of Bit Error for Coherently Detected BPSK

An important measure of performance used for comparing digital modulation schemes is the probability of error, P_E. For the correlator or matched filter, the calculations for obtaining P_E can be viewed geometrically. (See Figure 4.6.) They involve finding the probability that given a particular transmitted signal vector, say $\mathbf{s}_1$, the noise vector $\mathbf{n}$ will give rise to a received signal falling outside region 1. The probability of the detector making an incorrect decision is termed the *probability of symbol error* (P_E). It is often convenient to specify system performance by the probability of bit error (P_B), even when decisions are made on the basis of symbols

for which $M > 2$. The relationship between P_B and P_E is treated in Section 4.9.3 for orthogonal signaling and in Section 4.9.4 for multiple phase signaling.

For convenience, this section is restricted to the coherent detection of BPSK modulation. For this case the symbol error probability is the bit error probability. Assume that the signals are equally likely. Also assume that when signal $s_i(t)$ $(i = 1, 2)$ is transmitted, the received signal $r(t)$ is equal to $s_i(t) + n(t)$, where $n(t)$ is an AWGN process, and any degradation effects due to channel-induced ISI or circuit-induced ISI have been neglected. The antipodal signals $s_1(t)$ and $s_2(t)$ can be characterized in a one-dimensional signal space as described in Section 4.4.1, where

$$\left.\begin{aligned} s_1(t) &= \sqrt{E}\,\psi_1(t) \\ s_2(t) &= -\sqrt{E}\,\psi_1(t) \end{aligned}\right\} 0 \le t \le T \tag{4.74}$$

and

The detector will choose the $s_i(t)$ with the largest correlator output $z_i(T)$; or, in this case of equal-energy antipodal signals, the detector, using the decision rule in Equation (4.20), decides on the basis of

$$\begin{aligned} &s_1(t) \quad \text{if } z(T) > \gamma_0 = 0 \\ &s_2(t) \quad \text{otherwise} \end{aligned} \tag{4.75}$$

and

Two types of errors can be made, as shown in Figure 4.9: The first type of error takes place if signal $s_1(t)$ is transmitted but the noise is such that the detector measures a negative value for $z(T)$ and chooses hypothesis H_2, the hypothesis that signal $s_2(t)$ was sent. The second type of error takes place if signal $s_2(t)$ is transmitted but the detector measures a positive value for $z(T)$ and chooses hypothesis H_1, the hypothesis that signal $s_1(t)$ was sent.

In Section 3.2.1.1, an expression was developed in Equation (3.42) for the probability of a bit error P_B, for this binary *minimum error* detector. We rewrite this relationship as

$$P_B = \int_{(a_1 - a_2)/2\sigma_0}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du = Q\left(\frac{a_1 - a_2}{2\sigma_0}\right) \tag{4.76}$$

where σ_0 is the standard deviation of the noise out of the correlator. The function $Q(x)$, called the *complementary error function* or *co-error function.* is defined as

$$Q(X) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} \exp\left(-\frac{u^2}{2}\right) du \tag{4.77}$$

and is described in greater detail in Sections 3.2 and B.3.2.

For equal-energy antipodal signaling, such as the BPSK format in Equation (4.74), the receiver output signal components are $a_1 = \sqrt{E_b}$ when $s_1(t)$ is sent and $a_2 = -\sqrt{E_b}$ when $s_2(t)$ is sent, where E_b is the signal energy per binary symbol. For AWGN we can replace the noise variance σ_0^2 out of the correlator with $N_0/2$ (see Appendix C), so that we can rewrite Equation (4.76) as follows:

$$P_B = \int_{\sqrt{2E_b/N_0}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du \tag{4.78}$$

$$= Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{4.79}$$

This result for bandpass antipodal BPSK signaling is the same as the results that were developed earlier in Equation (3.70) for the matched-filter detection of antipodal signaling in general, and in Equation (3.76) for the matched-filter detection of baseband antipodal signaling in particular. This is an example of an *equivalence theorem,* described earlier. For linear systems the equivalence theorem establishes that the mathematics of detection is unaffected by a shift in frequency. Hence in this chapter, the use of matched filters or correlators in the detection of bandpass signals yields the same relationships as those developed for comparable signals at baseband.

Example 4.4 Bit Error Probability for BPSK Signaling

Find the bit error probability for a BPSK system with a bit rate of 1 Mbit/s. The received waveforms $s_1(t) = A \cos \omega_0 t$ and $s_2(t) = -A \cos \omega_0 t$, are coherently detected with a matched filter. The value of A is 10 mV. Assume that the single-sided noise power spectral density is $N_0 = 10^{-11}$ W/Hz and that signal power and energy per bit are normalized relative to a 1-Ω load.

Solution

$$A = \sqrt{\frac{2E_b}{T}} = 10^{-2}\,\text{V} \qquad T = \frac{1}{R} = 10^{-6}\,\text{s}$$

Thus,

$$E_b = \frac{A^2}{2} T = 5 \times 10^{-11}\,\text{J} \qquad \text{and} \qquad \sqrt{\frac{2E_b}{N_0}} = 3.16$$

Also,

$$P_B = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) = Q(3.16)$$

Using Table B.1 or Equation (3.44), we obtain

$$P_B = 8 \times 10^{-4}$$

4.7.2 Probability of Bit Error for Coherently Detected, Differentially Encoded Binary PSK

Channel waveforms sometimes experience inversion; for example, when using a coherent reference generated by a phase-locked loop, one may have phase ambiguity. If the carrier phase were reversed in a DPSK modulation application, what would be the effect on the message? The only effect would be an error in the bit during

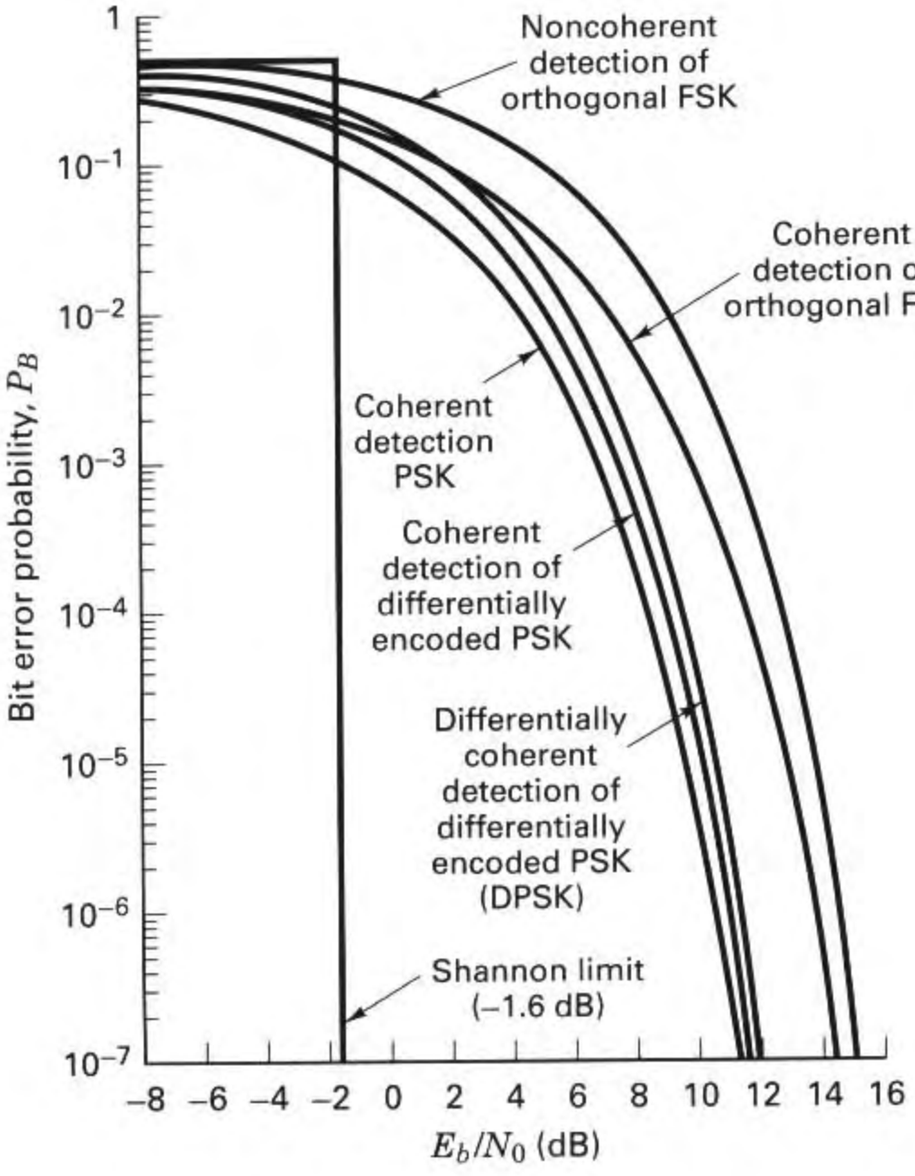


Figure 4.25 Bit error probability for several types of binary systems.

which inversion occurred or the bit just after inversion, since the message information is encoded in the similarity or difference between adjacent symbols. The similarity or difference quality remains unchanged if the carrier is inverted. Sometimes messages (and their assigned waveforms) are *differentially encoded* and *coherently detected* simply to avoid these phase ambiguities.

The probability of bit error for coherently detected, differentially encoded PSK is given by [5]

$$P_B = 2Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\left[1 - Q\left(\sqrt{\frac{2E_b}{N_0}}\right)\right] \tag{4.80}$$

This relationship is plotted in Figure 4.25. Notice that there is a small degradation of error performance compared with the coherent detection of PSK. This is due to the differential encoding, since any single detection error will usually result in two decision errors. Error performance for the more popular differentially coherent detection (DPSK) is covered in Section 4.7.5.

4.7.3 Probability of Bit Error for Coherently Detected Binary Orthogonal FSK

Equations (4.78) and (4.79) describe the probability of bit error for coherent antipodal signals. A more general treatment for binary coherent signals (not limited to antipodal signals) yields the following equation for P_B [6]:

$$P_B = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{(1-\rho)E_b/N_0}}^{\infty} \exp\left(-\frac{u^2}{2}\right) du \tag{4.81}$$

From Equation (3.64b), $\rho = \cos\theta$ is the time cross-correlation coefficient between signal $s_1(t)$ and $s_2(t)$, where θ is the angle between signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ (see Figure 4.6). For antipodal signals such as BPSK, $\theta = \pi$, thus $\rho = -1$.

For orthogonal signals such as binary FSK (BFSK), $\theta = \pi/2$, since the $\mathbf{s}_1$ and $\mathbf{s}_2$ vectors are perpendicular to each other; thus $\rho = 0$, as can be verified with Equation (3.64a), and Equation (4.81) can then be written as

$$P_B = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{E_b/N_0}}^{\infty} \exp\left(-\frac{u^2}{2}\right) du = Q\left(\sqrt{\frac{E_b}{N_0}}\right) \tag{4.82}$$

where the *co-error function* $Q(x)$ is described in greater detail in Sections 3.2 and B.3.2. The result in Equation (4.82) for the coherent detection of orthogonal BFSK plotted in Figure 4.25 is the same as the results that were developed earlier in Equation (3.71) for the matched-filter detection of orthogonal signaling, in general, and in Equation (3.73) for the matched-filter detection of baseband orthogonal signaling (unipolar pulses), in particular. The details of *on-off keying* (OOK) are not treated in this book. However, it is worth noting that on-off keying is an orthogonal signaling set (unipolar pulse signaling is the baseband equivalent of OOK). Thus the relationship in Equation (4.82) applies to the matched-filter detection of OOK, as it does to any coherent detection of orthogonal signaling.

The relationship in Equation (4.82) can also be confirmed by noting that the energy difference between the orthogonal signal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$, with amplitudes of $\sqrt{E_b}$, as shown in Figure 3.10b, can be computed as the square of the distance between the heads of the orthogonal vectors, which will be $E_d = 2E_b$. Using this result in Equation (3.63) also yields Equation (4.82). If we compare Equation (4.82) with Equation (4.79), we can see that 3-dB more E_b/N_0 is required for BFSK to provide the same performance as BPSK. It should not be surprising that the performance of BFSK signaling is 3-dB worse than BPSK signaling, since for a given signal power, the distance-squared between orthogonal vectors is a factor of two less than the distance squared between antipodal vectors.

4.7.4 Probability of Bit Error for Noncoherently Detected Binary Orthogonal FSK

Consider the equally likely binary orthogonal FSK signal set $\{s_1(t)\}$, defined in Equation (4.8) as follows:

$$s_i(t) = \sqrt{\frac{2E}{T}} \cos(\omega_i t + \phi) \qquad 0 \le t \le T, \qquad i = 1, 2$$

The phase term ϕ is unknown and assumed constant. The detector is characterized by $M = 2$ channels of bandpass filters and envelope detectors, as shown in Figure 4.19. The input to the detector consists of the received signal $r(t) = s_i(t) + n(t)$, where $n(t)$ is a white Gaussian noise process with two-sided power spectral density $N_0/2$. Assume that $s_1(t)$ and $s_2(t)$ are separated in frequency sufficiently that they have negligible overlap. For $s_1(t)$ and $s_2(t)$ being equally likely, we start the bit-error probability P_B computation with Equation (3.38) as we did for baseband signaling:

$$\begin{aligned} P_B &= \tfrac{1}{2} P(H_2|s_1) + \tfrac{1}{2} P(H_1|s_2) \\ &= \frac{1}{2} \int_{-\infty}^{0} p(z|s_1)\, dz + \frac{1}{2} \int_{0}^{\infty} p(z|s_2)\, dz \end{aligned} \tag{4.83}$$

For the binary case, the *test statistic* $z(T)$ is defined by $z_1(T) - z_2(T)$. Assume that the bandwidth of the filter W_f is $1/T$, so that the envelope of the FSK signal is (approximately) preserved at the filter output. If there was no noise at the receiver, the value of $z(T) = \sqrt{2E/T}$ when $s_1(t)$ is sent, and $z(T) = -\sqrt{2E/T}$ when $s_2(t)$ is sent. Because of this symmetry, the optimum threshold is $\gamma_0 = 0$. The pdf $p(z|s_1)$ is similar to $p(z|s_2)$; that is,

$$p(z|s_1) = p(-z|s_2) \tag{4.84}$$

Therefore, we can write

$$P_B = \int_{0}^{\infty} p(z|s_2)\, dz \tag{4.85}$$

or

$$P_B = P(z_1 > z_2|s_2) \tag{4.86}$$

where z_1 and z_2 denote the outputs $z_1(T)$ and $z_2(T)$ from the envelope detectors shown in Figure 4.19. For the case in which the tone $s_2(t) = \cos \omega_2 t$ is sent, such that $r(t) = s_2(t) + n(t)$, the output $z_1(T)$ is a *Gaussian noise random variable only;* it has no signal component. A Gaussian distribution into the *nonlinear envelope detector* yields a Rayleigh distribution at the output [6], so that

$$p(z_1|s_2) = \begin{cases} \dfrac{z_1}{\sigma_0^2} \exp\left(-\dfrac{z_1^2}{2\sigma_0^2}\right) & z_1 \ge 0 \\ 0 & z_1 < 0 \end{cases} \tag{4.87}$$

where σ_0^2 is the noise at the filter output. On the other hand, $z_2(T)$ has a Rician distribution, since the input to the lower envelope detector is a sinusoid plus noise [6]. The pdf $p(z_2|s_2)$ is written as

$$p(z_2|s_2) = \begin{cases} \dfrac{z_2}{\sigma_0^2} \exp\left[-\dfrac{(z_2^2 + A^2)}{2\sigma_0^2}\right] I_0\left(\dfrac{z_2 A}{\sigma_0^2}\right) & z_2 \geq 0 \\ 0 & z_2 < 0 \end{cases} \tag{4.88}$$

where $A = \sqrt{2E/T}$, and as before, σ_0^2 is the noise at the filter output. The function $I_0(x)$, known as the modified zero-order Bessel function of the first kind [7], is defined as

$$I_0(x) = \frac{1}{2\pi} \int_0^{2\pi} \exp(x \cos\theta)\, d\theta \tag{4.89}$$

When $s_2(t)$ is transmitted, the receiver makes an error whenever the envelope sample $z_1(T)$ obtained from the upper channel (due to noise alone) exceeds the envelope sample $z_2(T)$ obtained from the lower channel (due to signal plus noise). Thus the probability of this error can be obtained by integrating $p(z_1|s_2)$ with respect to z_1 from z_2 to infinity, and then averaging over all possible values of z_2. That is,

$$P_B = P(z_1 > z_2|s_2)$$

$$= \int_0^\infty p(z_2|s_2) \left[\int_{z_2}^\infty p(z_1|s_2)\, dz_1 \right] dz_2 \tag{4.90}$$

$$= \int_0^\infty \frac{z_2}{\sigma_0^2} \exp\left[-\frac{(z_2^2 + A^2)}{2\sigma_0^2}\right] I_0\left(\frac{z_2 A}{\sigma_0^2}\right) \left[\int_{z_2}^\infty \frac{z_1}{\sigma_0^2} \exp\left(-\frac{z_1^2}{2\sigma_0^2}\right) dz_1 \right] dz_2 \tag{4.91}$$

where $A = \sqrt{2E/T}$ and where the inner integral is the conditioned probability of an error for a fixed value of z_2, given that $s_2(t)$ was sent, and the outer integral averages this conditional probability over all possible values of z_2. This integral can be evaluated [8], yielding

$$P_B = \frac{1}{2} \exp\left(-\frac{A^2}{4\sigma_0^2}\right) \tag{4.92}$$

Using Equation (1.19), we can express the filter output noise as

$$\sigma_0^2 = 2\left(\frac{N_0}{2}\right) W_f \tag{4.93}$$

where $G_n(f) = N_0/2$ and W_f is the filter bandwidth. Thus Equation (4.92) becomes

$$P_B = \frac{1}{2} \exp\left(-\frac{A^2}{4N_0 W_f}\right) \tag{4.94}$$

Equation (4.94) indicates that the error performance depends on the bandpass filter bandwidth, and that P_B becomes smaller as W_f is decreased. The result is valid only when the intersymbol interference (ISI) is negligible. The minimum W_f allowed (i.e., for no ISI) is obtained from Equation (3.81) with the filter roll-off factor $r = 0$. Thus $W_f = R$ bits/s $= 1/T$, and we can write Equation (4.94) as

$$P_B = \frac{1}{2}\exp\left(-\frac{A^2T}{4N_0}\right) \tag{4.95}$$

$$= \frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right) \tag{4.96}$$

where $E_b = (1/2)A^2T$ is the energy per bit. When comparing the error performance of noncoherent FSK with coherent FSK (see Figure 4.25), it is seen that for the same P_B, noncoherent FSK requires approximately 1 dB more E_b/N_0 than that for coherent FSK (for $P_B \leq 10^{-4}$). The noncoherent receiver is easier to implement, because coherent reference signals need not be generated. Therefore, almost all FSK receivers use noncoherent detection. It can be seen in the following section that when comparing noncoherent orthogonal FSK to noncoherent DPSK, the same 3-dB difference occurs as for the comparison between coherent orthogonal FSK and coherent PSK.

As mentioned earlier, the details of on-off keying (OOK) are not treated in this book. However, it is worth noting that the bit error probability P_B described in Equation (4.96) is identical to the P_B for the noncoherent detection of OOK signaling.

4.7.5 Probability of Bit Error for Binary DPSK

Let us define a BPSK signal set as follows:

$$\begin{aligned} x_1(t) &= \sqrt{\frac{2E}{T}}\cos(\omega_0 t + \phi) \qquad 0 \leq t \leq T \\ x_2(t) &= \sqrt{\frac{2E}{T}}\cos(\omega_0 t + \phi \pm \pi) \qquad 0 \leq t \leq T \end{aligned} \tag{4.97}$$

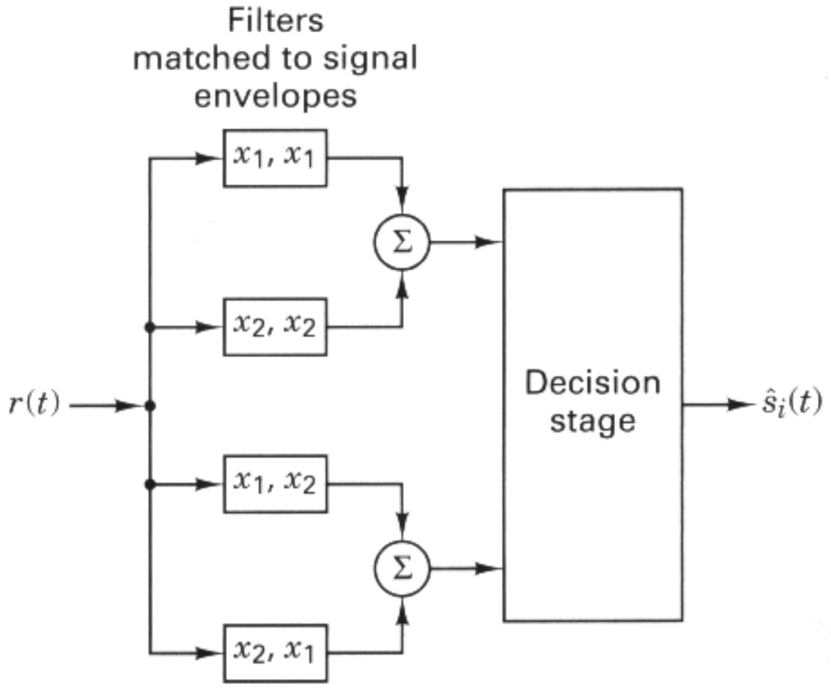
A characteristic of DPSK is that there are no fixed decision regions in the signal space. Instead, the decision is based on the phase difference between successively received signals. Then for DPSK signaling we are really transmitting each bit with the binary signal pair

$$\begin{aligned} &\text{and} \qquad s_1(t) = (x_1, x_1) \quad \text{or} \quad (x_2, x_2) \qquad 0 \leq t \leq 2T \\ &\qquad\quad\ s_2(t) = (x_1, x_2) \quad \text{or} \quad (x_2, x_1) \qquad 0 \leq t \leq 2T \end{aligned} \tag{4.98}$$

where (x_i, x_j) $(i, j = 1, 2)$ denotes $x_i(t)$ followed by $x_j(t)$ defined in Equation (4.97). The first T seconds of each waveform are actually the last T seconds of the previous waveform. Note that $s_1(t)$ and $s_2(t)$ can each have either of two possible forms and that $x_1(t)$ and $x_2(t)$ are antipodal signals. Thus the correlation between $s_1(t)$ and $s_2(t)$ for *any combination* of forms can be written as

$$z(2T) = \int_0^{2T} s_1(t)s_2(t)\,dt$$
$$= \int_0^{T} [x_1(t)]^2\,dt - \int_0^{T} [x_1(t)]^2\,dt = 0 \tag{4.99}$$

Therefore, pairs of DPSK signals can be represented as orthogonal signals $2T$ seconds long. Detection could correspond to noncoherent envelope detection with four channels matched to each of the possible envelope outputs, as shown in Figure 4.26a. Since the two envelope detectors representing each symbol are negatives of each other, the envelope sample of each will be the same. Hence we can implement the detector as a single channel for $s_1(t)$ matched to either (x_1, x_1) or (x_2, x_2), and a single channel for $s_2(t)$ matched to either (x_1, x_2) or (x_2, x_1), as shown in Figure 4.26b. The DPSK detector is therefore reduced to a standard two-channel noncoherent detector. In reality, the filter can be matched to the difference signal so that only one channel is necessary. In Figure 4.26, the filters are matched to the signal envelopes (over two symbol times). What does this mean in light of the fact that



(a)

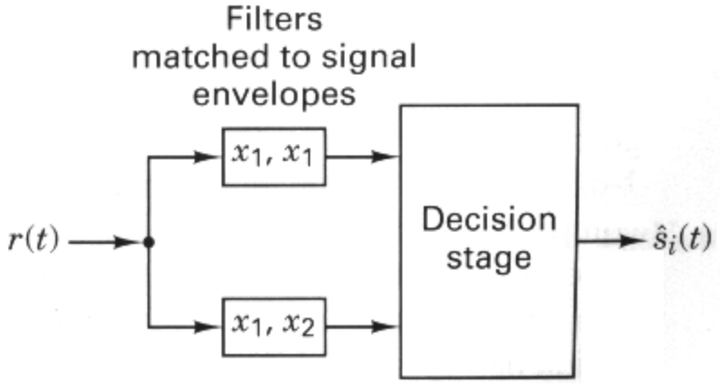


Figure 4.26 DPSK detection. (a) Four-channel differentially coherent detection of binary DPSK. (b) Equivalent two-channel detector for binary DPSK.

DPSK is a constant envelope signaling scheme? It means that we need to implement an energy detector, similar to the quadrature receiver in Figure 4.18, with each of the I and Q reference signals occuring in pairs during the time interval $(0 \le t \le 2T)$, as follows:

$$\begin{aligned} s_1(t)\ I \text{ reference:} &\quad \sqrt{2/T}\cos\omega_0 t,\ \sqrt{2/T}\cos\omega_0 t \\ s_1(t)\ Q \text{ reference:} &\quad \sqrt{2/T}\sin\omega_0 t,\ \sqrt{2/T}\sin\omega_0 t \end{aligned}$$

$$\begin{aligned} s_2(t)\ I \text{ reference:} &\quad \sqrt{2/T}\cos\omega_0 t,\ -\sqrt{2/T}\cos\omega_0 t \\ s_2(t)\ Q \text{ reference:} &\quad \sqrt{2/T}\sin\omega_0 t,\ -\sqrt{2/T}\sin\omega_0 t \end{aligned}$$

Since pairs of DPSK signals are orthogonal, such noncoherent detection operates with the bit-error probability given by Equation (4.96). However, since the DPSK signals have a bit interval of $2T$, the $s_i(t)$ signals defined in Equation (4.98) have twice the energy of a signal defined over a single-symbol duration. Thus, we may write P_B as

$$P_B = \frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right) \tag{4.100}$$

Equation (4.100) is plotted in Figure 4.25, designated as differentially coherent detection of differentially encoded PSK, or simply DPSK. This expression is valid for the optimum DPSK detector shown in Figure 4.17c. For the detector shown in Figure (4.17b), the error probability will be slightly inferior to that given in Equation (4.100) [3]. When comparing the error performance of Equation (4.100) with that of coherent PSK (see Figure 4.25), it is seen that for the same P_B, DPSK requires approximately 1 dB more E_b/N_0 than does BPSK (for $P_B \le 10^{-4}$). It is easier to implement a DPSK system than a PSK system, since the DPSK receiver does not need phase synchronization. For this reason, DPSK, although less efficient than PSK, is sometimes the preferred choice between the two.

4.7.6 Comparison of Bit Error Performance for Various Modulation Types

The P_B expressions for the best known of the binary modulation schemes discussed above are listed in Table 4.1 and are illustrated in Figure 4.25. For $P_B = 10^{-4}$, it can be seen that there is a difference of approximately 4-dB between the best (coherent PSK) and the worst (noncoherent orthogonal FSK) that were discussed here. In some cases, 4 dB is a small price to pay for the implementation simplicity gained in going from coherent PSK to noncoherent FSK; however, for other cases, even a 1-dB saving is worthwhile. There are other considerations besides P_B and system complexity; for example, in some cases (such as a randomly fading channel), a noncoherent system is more desirable because there may be difficulty in establishing and maintaining a coherent reference. Signals that can withstand significant degradation before their ability to be detected is affected are clearly desirable in military and space applications.

TABLE 4.1 Probability of Error for Selected Binary Modulation Schemes

Modulation	P_B
PSK (coherent)	$Q\left(\sqrt{\frac{2E_b}{N_0}}\right)$
DPSK (differentially coherent)	$\frac{1}{2}\exp\left(-\frac{E_b}{N_0}\right)$
Orthogonal FSK (coherent)	$Q\left(\sqrt{\frac{E_b}{N_0}}\right)$
Orthogonal FSK (noncoherent)	$\frac{1}{2}\exp\left(-\frac{1}{2}\frac{E_b}{N_0}\right)$

4.8 *M*-ARY SIGNALING AND PERFORMANCE

4.8.1 Ideal Probability of Bit Error Performance

The typical probability of error versus E_b/N_0 curve was shown to have a waterfall-like shape in Figure 3.6. The probability of bit error (P_B) characteristics of various binary modulation schemes in AWGN also display this shape, as shown in Figure 4.25. What should an *ideal* P_B versus E_b/N_0 curve look like? Figure 4.27 displays the ideal characteristic as the *Shannon limit.* The limit represents the threshold E_b/N_0 below which reliable communication cannot be maintained. Shannon's work is described in greater detail in Chapter 9.

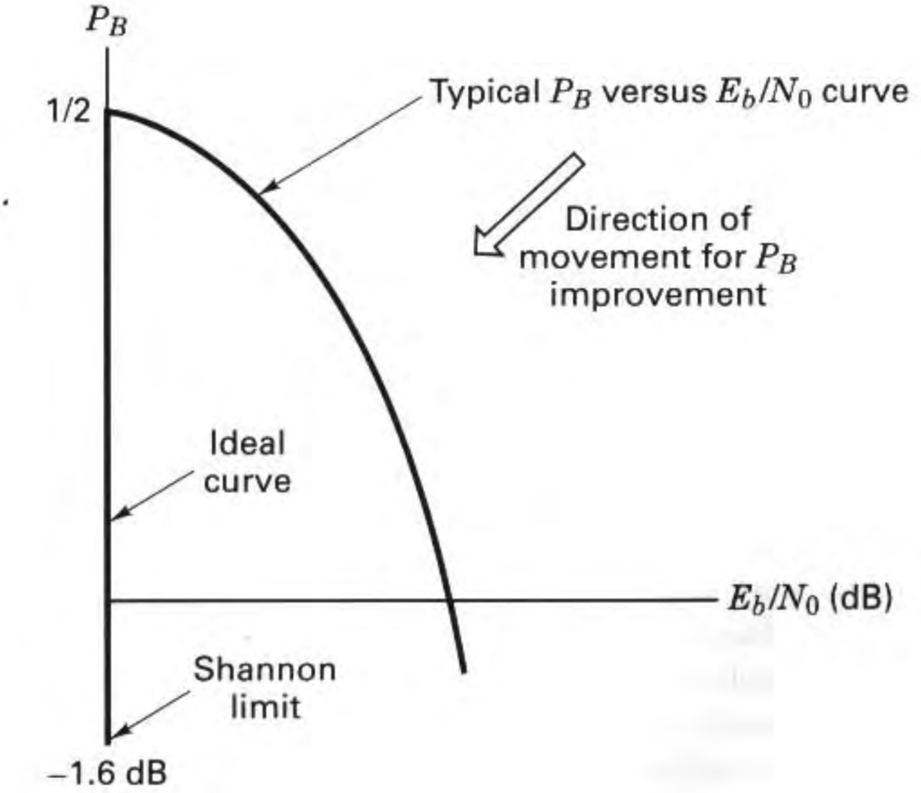


Figure 4.27 Ideal P_B versus E_b/N_0 curve.

We can describe the ideal curve in Figure 4.27 as follows. For all values of E_b/N_0 above the Shannon limit of −1.6 dB, P_B is zero. Once E_b/N_0 is reduced below the Shannon limit, P_B degrades to the worst-case value of $\frac{1}{2}$. (Note that $P_B = 1$ is not the worst case for binary signaling, since that value is just as good as $P_B = 0$; if the probability of making a bit error is 100%, the bit stream could simply be inverted to retrieve the correct data.) It should be clear, by comparing the typical P_B curve with the ideal one in Figure 4.27 that the large arrow in the figure describes the desired direction of movement to achieve improved P_B performance.

4.8.2 *M*-ary Signaling

Let us review M-ary signaling. The processor considers k bits at a time. It instructs the modulator to produce one of $M = 2^k$ waveforms; binary signaling is the special case where $k = 1$. Does M-ary signaling improve or degrade error performance? (Be careful with your answer—the question is a loaded one.) Figure 4.28 illustrates the probability of bit error $P_B(M)$ versus E_b/N_0 for coherently detected *orthogonal* M-ary signaling over a Gaussian channel. Figure 4.29 similarly illustrates $P_B(M)$ versus E_b/N_0 for coherently detected *multiple phase* M-ary signaling over a Gaussian channel. In which direction do the curves move as the value of k (or M) increases? From Figure 4.27 we know the directions of curve movement for improved and degraded error performance. In Figure 4.28, as k increases, the curves move in the direction of improved error performance. In Figure 4.29, as k increases, the curve move in the direction of degraded error performance. Such movement tells us that M-ary signaling produces improved error performance with orthogonal signaling and degraded error performance with multiple phase signaling. Can that be true? Why would anyone ever use multiple phase PSK signaling if it provides degraded error performance compared to binary PSK signaling? It *is* true, and many systems do use multiple phase signaling. The question, as stated, is loaded because it implies that error probability versus E_b/N_0 is the *only* performance criterion; there are many others (e.g., bandwidth, throughput, complexity, cost), but in Figures 4.28 and 4.29 error performance is the characteristic that stands out explicitly.

A performance characteristic that is not explicitly seen in Figures 4.28 and 4.29 is the required system bandwidth. For the curves characterizing M-ary orthogonal signals in Figure 4.28, as k increases, the required bandwidth also increases. For the M-ary multiple phase curves in Figure 4.29, as k increases, a larger bit rate can be transmitted within the same bandwidth. In other words, for a fixed data rate, the required bandwidth is decreased. Therefore, *both* the orthogonal and multiple phase error performance curves tell us that M-ary signaling represents a vehicle for performing a system trade-off. In the case of orthogonal signaling, error performance improvement can be achieved at the expense of bandwidth. In the case of multiple phase signaling, bandwidth performance can be achieved at the expense of error performance. Error performance versus bandwidth performance, a fundamental communications trade-off, is treated in greater detail in Chapter 9.

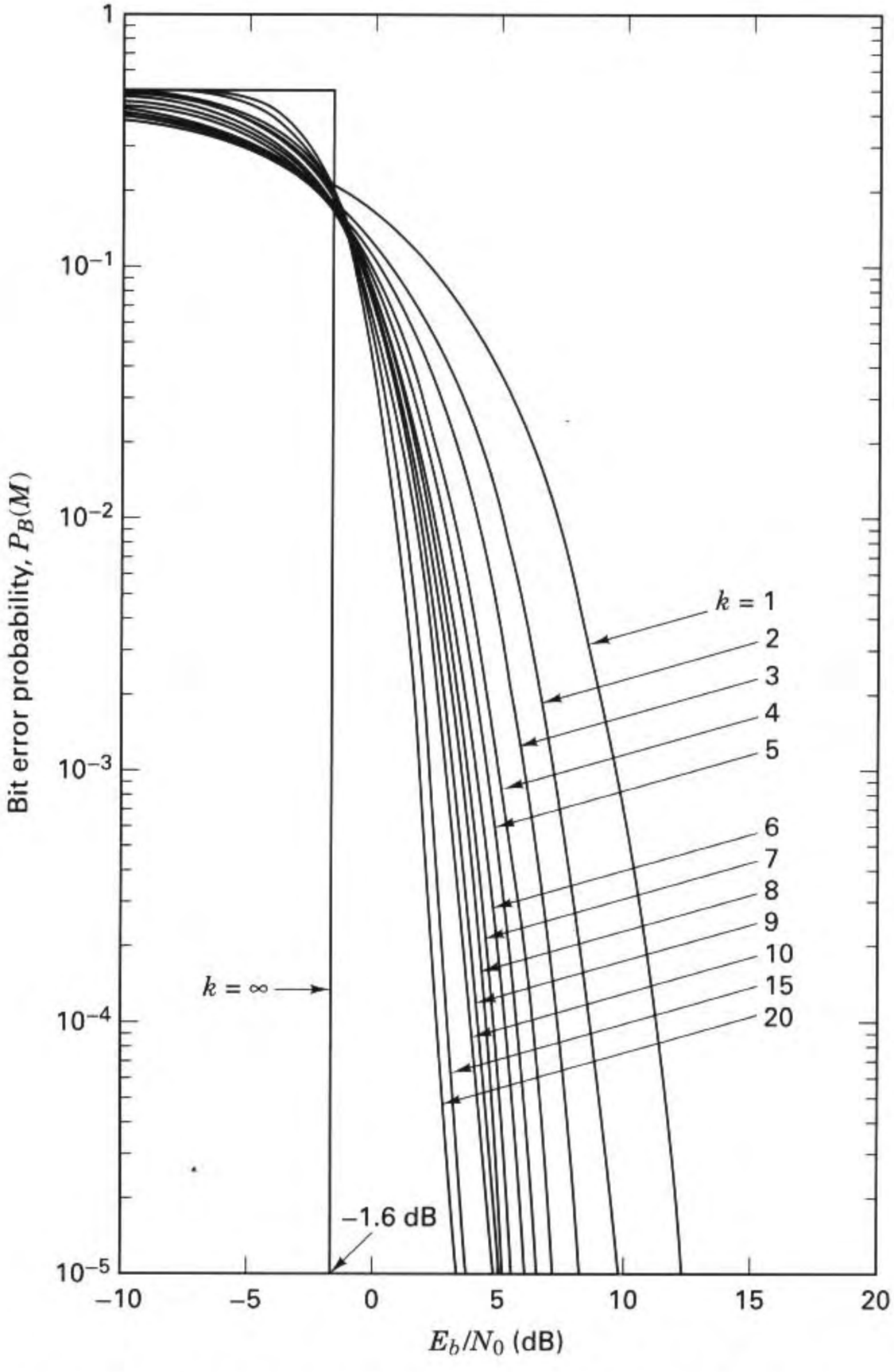


Figure 4.28 Bit error probability for coherently detected *M*-ary orthogonal signaling. (Reprinted from W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering,* Prentice Hall, Inc., Englewood Cliffs, N.J., 1973, courtesy of W. C. Lindsey and Marvin K. Simon).

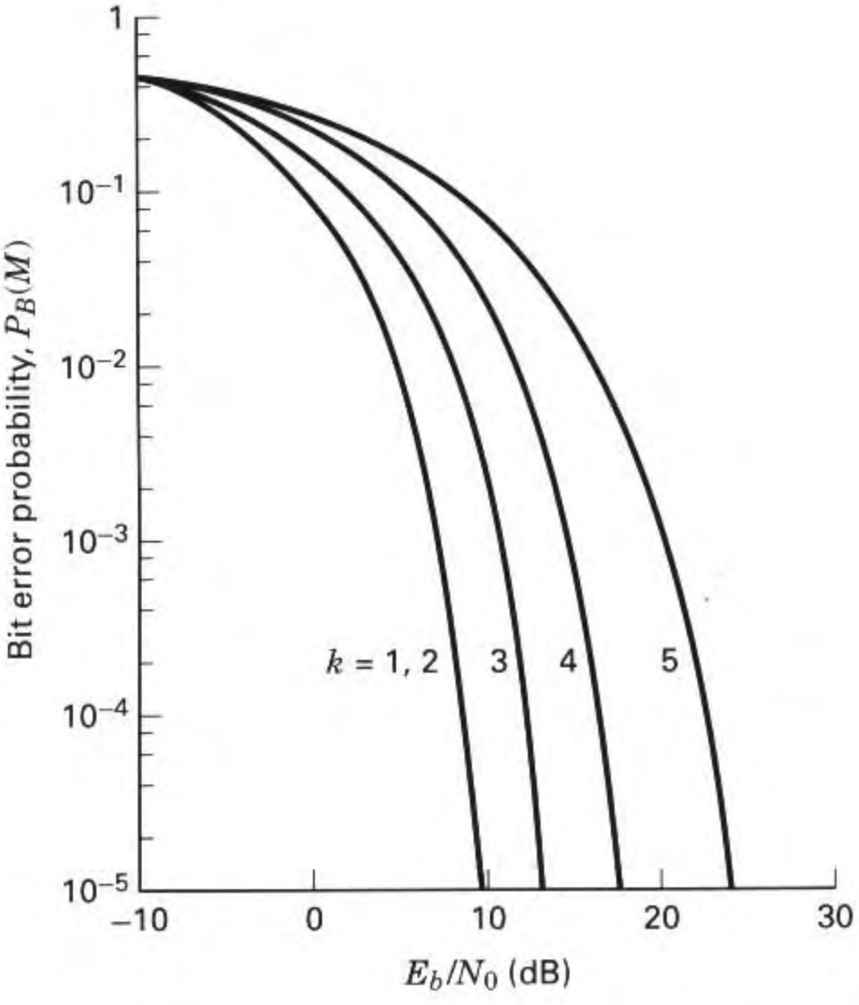


Figure 4.29 Bit error probability for coherently detected multiple phase signaling.

4.8.3 Vectorial View of MPSK Signaling

Figure 4.30 illustrates MPSK signal sets for $M = 2, 4, 8$, and 16. In Figure 4.30a we see the binary ($k = 2$, $M = 2$) antipodal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ positioned 180° apart. The decision boundary is drawn so as to partition the signal space into two regions. On the figure is also shown a noise vector $\mathbf{n}$ equal in magnitude to $\mathbf{s}_1$. The figure establishes the magnitude and orientation of the minimum energy noise vector that would cause the detector to make a symbol error.

In Figure 4.30b we see the 4-ary ($k = 2$, $M = 4$) vectors positioned 90° apart. The decision boundaries (only one line is drawn) divide the signal space into four regions. Again a noise vector $\mathbf{n}$ is drawn (from the head of a signal vector, normal to the closest decision boundary) to illustrate the minimum energy noise vector that would cause the detector to make a symbol error. Notice that the minimum energy noise vector of Figure 4.30b is smaller than that of Figure 4.30a, illustrating that the 4-ary system is more vulnerable to noise than the 2-ary system (signal energy being equal for each case). As we move on to Figure 4.30c for the 8-ary case and Figure 4.30d for the 16-ary case, it should be clear that for multiple phase signaling, as M increases, we are crowding more signal vectors into the signal plane. As the vectors are moved closer together, a smaller amount of noise energy is required to cause an error.

Figure 4.30 adds some insight as to why the curves of Figure 4.29 behave as they do as k is increased. Figure 4.30 also provides some insight into a basic trade-

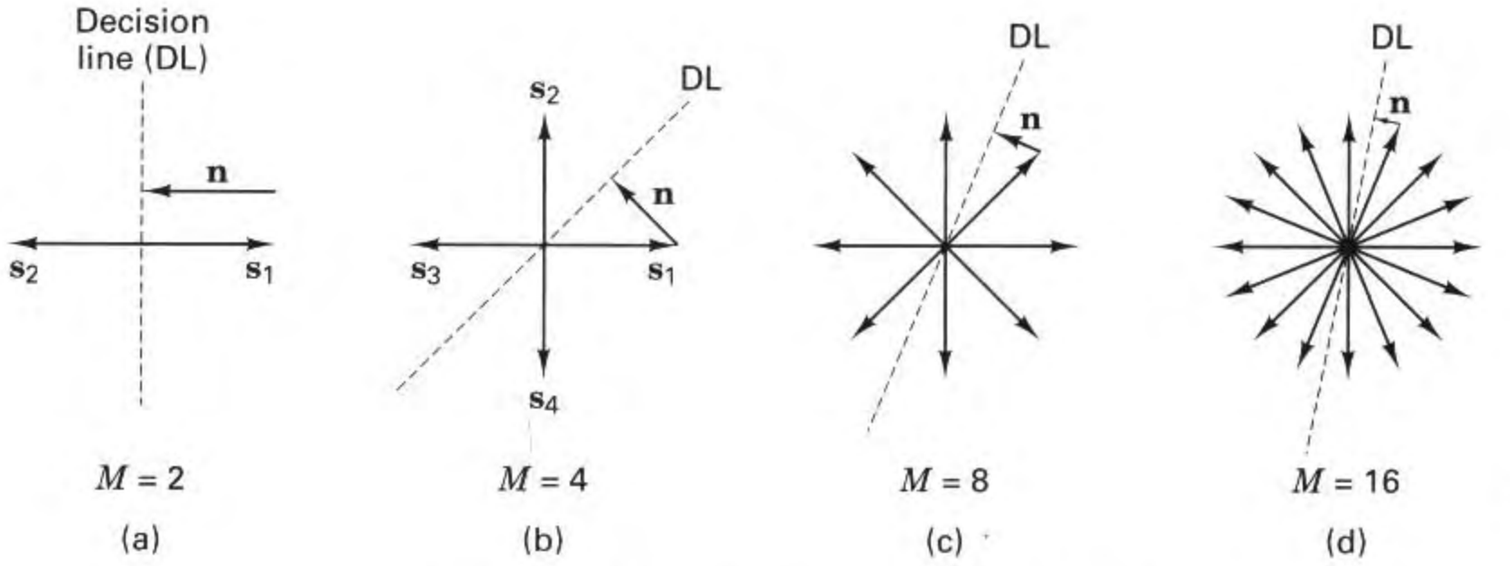


Figure 4.30 MPSK signal sets for $M = 2, 4, 8, 16$.

off in multiple phase signaling. Crowding more signal vectors into the signal space is tantamount to increasing the data rate without increasing the system bandwidth (the vectors are all confined to the same plane). In other words, we have increased the bandwidth utilization at the expense of error performance. Look at Figure 4.30d, where the error performance is worse than any of the other examples in Figure 4.30. How might we "buy back" the degraded error performance? In other words, what can we trade-off so that the distance between neighboring signal vectors in Figure 4.30d is increased to that in Figure 4.30a? We can increase the signal strength (make the signal vectors larger) until the minimum distance from the head of a signal vector to a decision line equals the length of the noise vector in Figure 4.30a. Therefore, in a multiple phase system, as M is increased, we can either achieve improved bandwidth performance at the expense of error performance, or if we increase the E_b/N_0 so that the error probability is not degraded, we can achieve improved bandwidth performance at the expense of increasing E_b/N_0.

Note that Figure 4.30 has been sketched so that all phasors have the same length for any of the M-ary cases. This is tantamount to saying that the comparisons are being considered for a fixed E_s/N_0, where E_s is symbol energy. The figure can also be drawn for a fixed E_b/N_0, in which case the phasor magnitudes would increase with increasing M. The phasors for $M = 4$, 8, and 16 would then have lengths greater than the $M = 2$ case by the factors $\sqrt{2}$, $\sqrt{3}$, and 2 respectively. We would still see crowding and increased vulnerability to noise, with increasing M, but the appearance would not be as pronounced as it is in Figure 4.30.

4.8.4 BPSK and QPSK Have the Same Bit Error Probability

In Equation (3.30) we stated the general relationship between E_b/N_0 and S/N which is rewritten

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{W}{R}\right) \tag{4.101}$$

where S is the average signal power and R is the bit rate. A BPSK signal with the available E_b/N_0 found from Equation (4.101) will perform with a P_B that can be read from the $k = 1$ curve in Figure 4.29. QPSK can be characterized as two orthogonal BPSK channels. The QPSK bit stream is usually partitioned into an even and odd (I and Q) stream; each new stream modulates an orthogonal component of the carrier at half the bit rate of the original stream. The I stream modulates the $\cos \omega_0 t$ term and the Q stream modulates the $\sin \omega_0 t$ term. If the magnitude of the original QPSK vector has the value A, the magnitude of the I and Q component vectors each has a value of $A/\sqrt{2}$, as shown in Figure 4.31. Thus, each of the quadrature BPSK signals has half of the average power of the original QPSK signal. Hence if the original QPSK waveform has a bit rate of R bits/s and an average power of S watts, the quadrature partitioning results in each of the BPSK waveforms having a bit rate of $R/2$ bits/s and an average power of $S/2$ watts.

Therefore, the E_b/N_0 characterizing each of the orthogonal BPSK channels, making up the QPSK signal, is equivalent to the E_b/N_0 in Equation (4.101), since it can be written as

$$\frac{E_b}{N_0} = \frac{S/2}{N_0}\left(\frac{W}{R/2}\right) = \frac{S}{N_0}\left(\frac{W}{R}\right) \tag{4.102}$$

Thus each of the orthogonal BPSK channels, and hence the composite QPSK signal, is characterized by the same E_b/N_0 and hence the same P_B performance as a BPSK signal. The natural orthogonality of the 90° phase shifts between adjacent QPSK symbols results in the *bit error probabilities* being equal for both BPSK and QPSK signaling. It is important to note that the symbol *error probabilities* are *not* equal for BPSK and QPSK signaling. The relationship between bit error probability and symbol error probability is treated in Sections 4.9.3 and 4.9.4. We see that, in effect, QPSK is the equivalent of two BPSK channels in quadrature. This same idea can be extended to any symmetrical M-ary amplitude/phase signaling, such as quadrature amplitude modulation (QAM) described in Section 9.8.3.

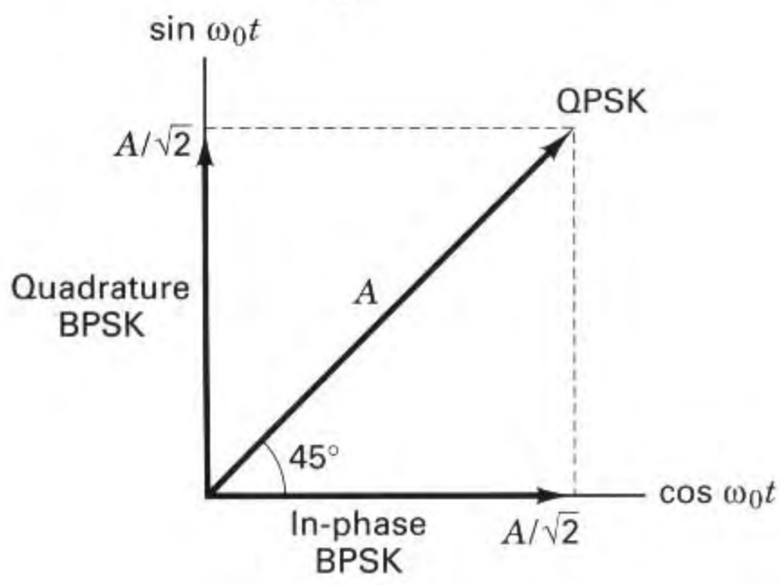


Figure 4.31 In-phase and quadrature BPSK components of QPSK signaling.

4.8.5 Vectorial View of MFSK Signaling

In Section 4.8.3, Figure 4.30 provides some insight as to why the error performance of MPSK signaling degrades as k (or M) increases. It would be useful to have a similar vectorial illustration for the error performance of orthogonal MFSK signaling as seen in the curves of Figure (4.28). Since the MFSK signal space is characterized by M mutually perpendicular axes, we can only conveniently illustrate the cases $M = 2$ and $M = 3$. In Figure 4.32a we see the binary orthogonal vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ positioned 90° apart. The decision boundary is drawn so as to partition the signal space into two regions. On the figure is also shown a noise vector $\mathbf{n}$, which represents the minimum noise vector that would cause the detector to make an error.

In Figure 4.32b we see a 3-ary signal space with axes positioned 90° apart. Here decision planes partition the signal space into three regions. Noise vectors $\mathbf{n}$ are shown added to each of the prototype signal vectors $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$; each noise vector illustrates an example of the minimum noise vector that would cause the detector to make a symbol error. The minimum noise vectors in Figure 4.32b are the same length as the noise vector in Figure 4.32a. In Section 4.4.4 we stated that for a given level of received energy, the distance between any two prototype signal vectors $\mathbf{s}_i$ and $\mathbf{s}_j$ in an M-ary orthogonal space is constant. It follows that the minimum distance between a prototype signal vector and any of the decision boundaries remains fixed as M increases. Unlike the case of MPSK signaling, where adding new signals to the signal set makes the signals vulnerable to smaller noise vectors, here, in the case of MFSK signaling, adding new signals to the signal set does *not* make the signals vulnerable to smaller noise vectors.

It would be convenient to illustrate the point by drawing higher dimensional orthogonal spaces, but of course this is not possible. We can only use our "mind's eye" to understand that increasing the signal set M—by adding additional axes, where each new axis is mutually perpendicular to all the others—does not crowd

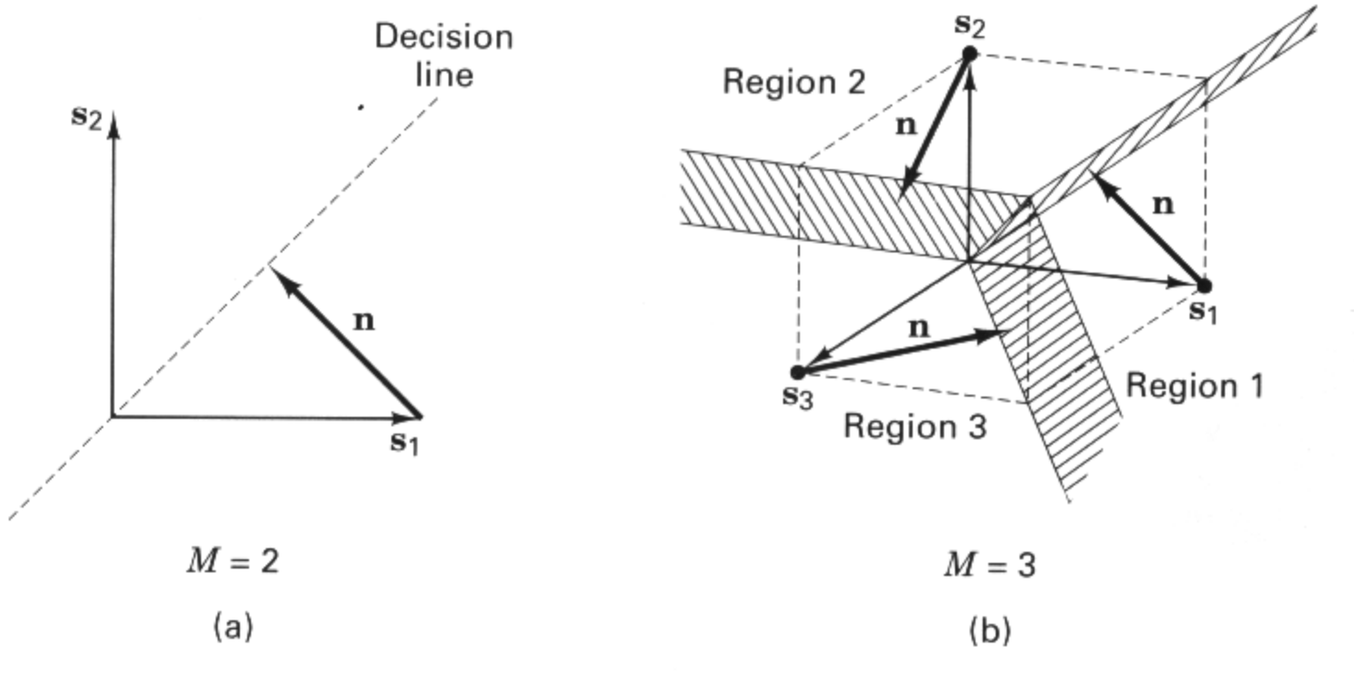


Figure 4.32 MFSK signal sets for $M = 2, 3$.

the signal set more closely together. Thus, a transmitted signal from an orthogonal set is *not* more vulnerable to a noise vector when the set is increased in size. In fact, we see from Figure 4.28 that as k increases, the bit error performance improves.

Understanding the error performance improvement of orthogonal signaling, as illustrated in Figure 4.28, is facilitated by comparing the probability of symbol error (P_E) versus unnormalized SNR, with P_E versus E_b/N_0. Figure 4.33 represents a set of P_E performance curves plotted against unnormalized SNR for coherent FSK signaling. Here we see that P_E degrades as M is increased. Didn't we say that a signal from an orthogonal set is *not* made more vulnerable to a given noise vector, as the orthogonal set is increased in size? It is correct that for orthogonal signaling, with a given SNR it takes a fixed size noise vector to perturb a transmitted signal into an error region; the signals do not become vulnerable to smaller noise vectors as M increases. However, as M increases, more neighboring decision regions are introduced; thus the number of ways in which a symbol error can be made increases. Figure 4.33 reflects the degradation in P_E versus unnormalized SNR as M is increased; there are $(M - 1)$ ways to make an error. Examining performance under the condition of a fixed SNR (as M increases) is not very useful for digital communications. A fixed SNR means a fixed amount of energy per symbol; thus as M increases, there is a fixed amount of energy to be apportioned over a larger number of bits, or there is less energy per bit. The most useful way of comparing one digital system with another is on the basis of *bit-normalized SNR* or E_b/N_0. The error performance improvement with increasing M (see Figure 4.28) manifests itself only when error probability is plotted against E_b/N_0. For this case, as M increases, the required E_b/N_0 (to meet a given error probability) is reduced for a fixed SNR; therefore, we need to map the plot shown in Figure 4.33 into a new plot, similar to that shown in Figure 4.28, where the abscissa represents E_b/N_0 instead of SNR. Figure 4.34 illustrates such a mapping; it demonstrates that curves manifesting degraded P_E with increasing M (such as Figure 4.33) are transformed into curves manifesting improved P_E with increasing M. The basic mapping relationship is expressed in Equation (4.101), repeated here as

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{W}{R}\right)$$

where W is the detection bandwidth. Since

$$R = \frac{\log_2 M}{T} = \frac{k}{T}$$

where T is the symbol duration, we can then write

$$\frac{E_b}{N_0} = \frac{S}{N}\left(\frac{WT}{\log_2 M}\right) = \frac{S}{N}\left(\frac{WT}{k}\right) \tag{4.103}$$

For FSK signaling, the detection bandwidth W (in hertz) is typically equal in value to the symbol rate $1/T$; in other words, $WT \approx 1$. Therefore,

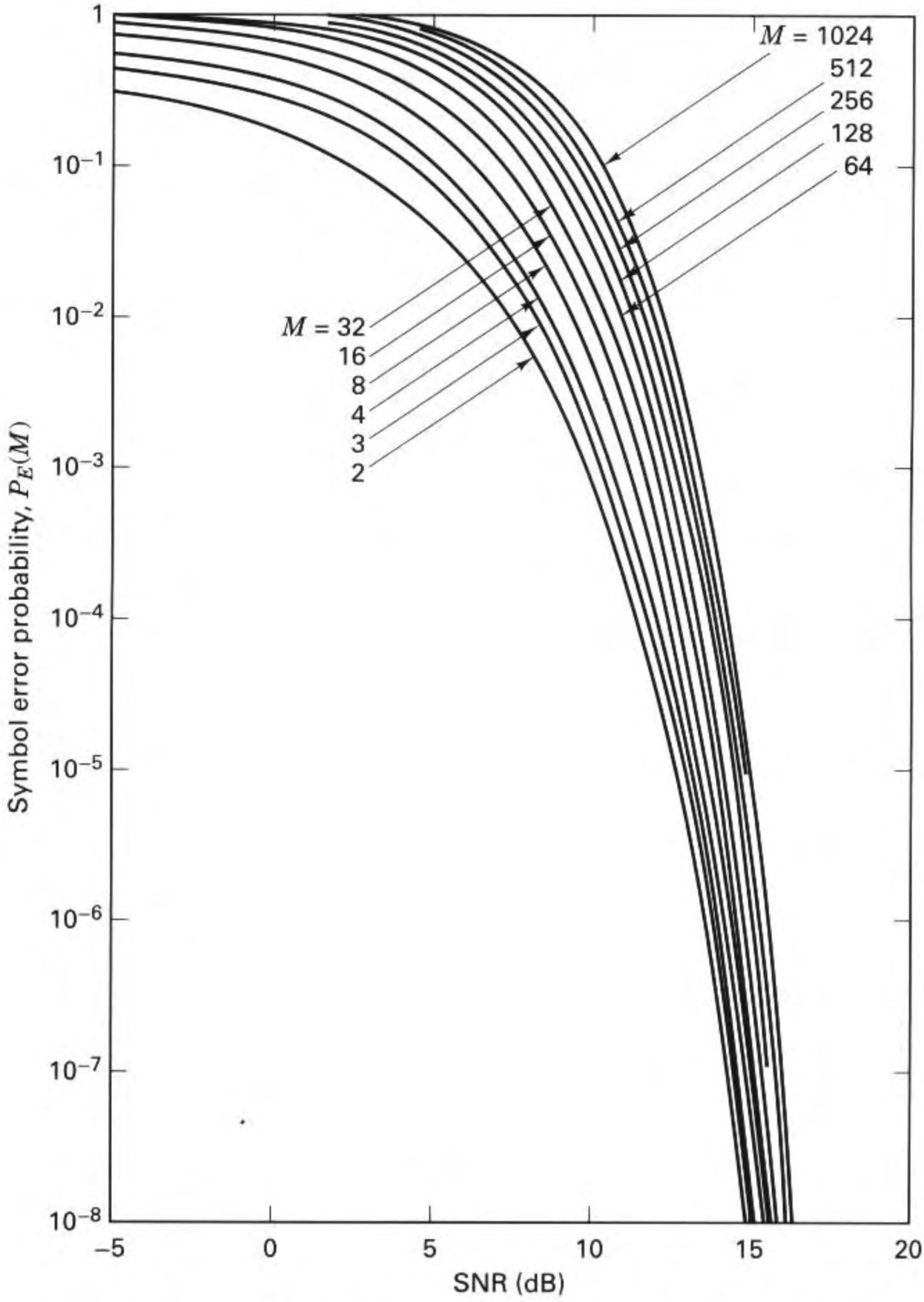


Figure 4.33 Symbol error probability versus SNR for coherent FSK signaling. (From Bureau of Standards, *Technical Note 167,* March 1963.) (Reprinted from *Central Radio Propagation Laboratory Technical Note 167,* March 25, 1963, Fig. 1, p. 5, courtesy of National Bureau of Standards.)

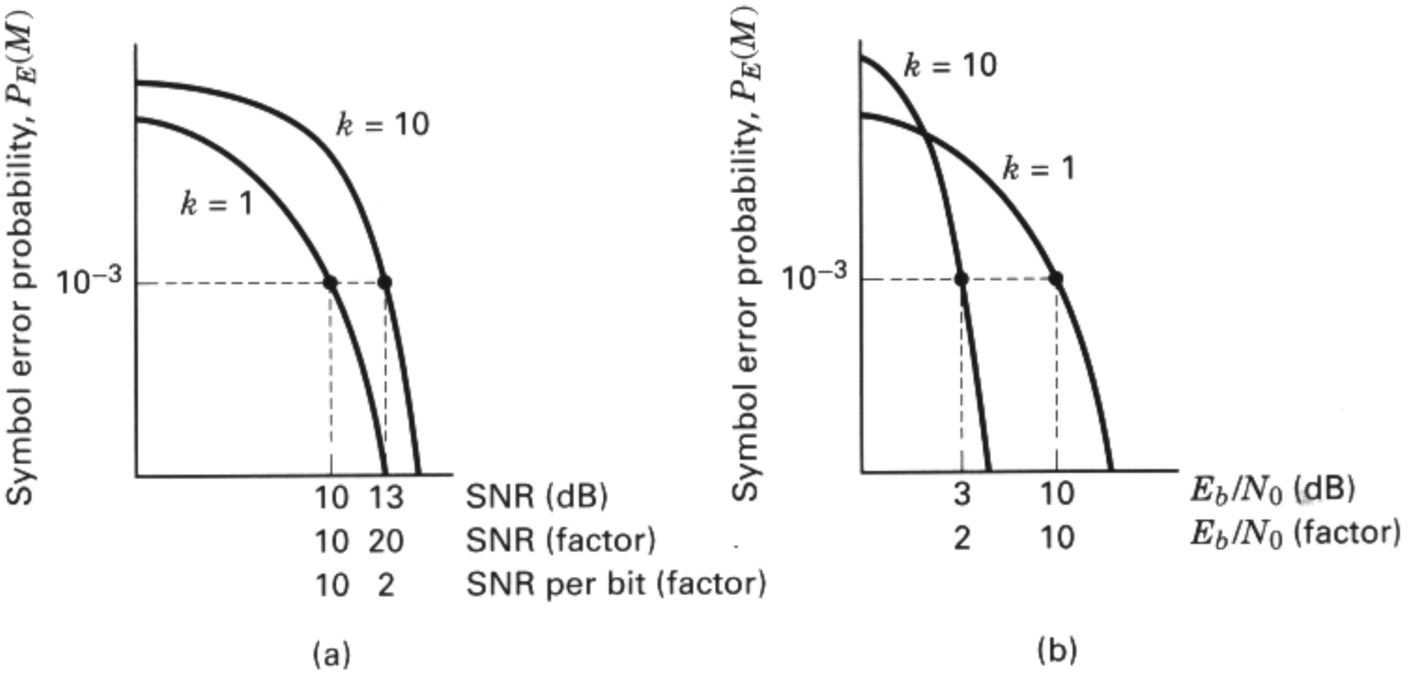


Figure 4.34 Mapping P_E versus SNR into P_E versus E_b/N_0 for orthogonal signaling. (a) Unnormalized. (b) Normalized.

$$\frac{E_b}{N_0} \approx \frac{S}{N}\left(\frac{1}{k}\right) \tag{4.104}$$

Figure 4.34 illustrates the mapping from P_E versus SNR to P_E versus E_b/N_0 for coherently detected M-ary orthogonal signaling, with "ballpark" numbers on the axes. In Figure 4.34a, on the $k = 1$ curve is shown an operating point corresponding to $P_E = 10^{-3}$ and SNR = 10 dB. On the $k = 10$ curve is shown an operating point at the same $P_E = 10^{-3}$ but with SNR = 13 dB (approximate values taken from Figure 4.33). Here we clearly see the degradation in error performance as k increases. To appreciate where the performance improvement comes from, let us convert the abscissa from the nonlinear scale of SNR in decibels to a linear one—SNR expressed as a factor. This is shown in Figure 4.34a as the factors 10 and 20 for the $k = 1$ and $k = 10$ cases, respectively. Next, we further convert the abscissa scale to SNR per bit (expressed as a factor). This is shown in Figure 4.34a as the factors 10 and 2 for the $k = 1$ and $k = 10$ cases, respectively. It is convenient to think of the 1024-ary symbol or waveform ($k = 10$ case) as being interchangeable with its 10-bit meaning. Thus, if the symbol requires 20 units of SNR then the 10 bits belonging to that symbol require that same 20 units; or, in other words, each bit requires 2 units.

Rather than performing such computations, we can simply map these same $k = 1$ and $k = 10$ cases onto the Figure 4.34b plane, representing P_E versus E_b/N_0. The $k = 1$ case looks exactly the same as it does in Figure 4.34a. But for the $k = 10$ case, there is a dramatic change. We can immediately see that signaling with the $k =$ 10-bit symbol requires only 2 units (3 dB) of E_b/N_0 compared with 10 units (10 dB) for the binary symbol. The mapping that gives rise to the required E_b/N_0 for the $k =$ 10 case is obtained from Equation (4.104) as follows: $E_b/N_0 = 20\ (1/10) = 2$ (or 3-dB), which shows the error performance improvement as k is increased. In digital communication systems, error performance is almost always considered in terms of

E_b/N_0, since such a measurement makes for a meaningful comparison between one system's performance and another. Therefore, the curves shown in Figures 4.33 and 4.34a are hardly ever seen.

Although Figure 4.33 is not often seen, we can still use it for gaining insight into why orthogonal signaling provides improved error performance as M or k increases. Let us consider the analogy of purchasing a commodity—say, grade A cottage cheese. The choice of the grade corresponds to some point on the P_E axis of Figure 4.33—say, 10^{-3}. From this point, construct a horizontal line through all of the curves (from $M = 2$ through $M = 1024$). At the grocery store we buy the very smallest container of cottage cheese, containing 2 ounces and costing \$1. On Figure 4.33 we can say that this purchase corresponds to our horizontal construct intercepting the $M = 2$ curve. We look down at the corresponding SNR and call the intercept on this axis our cost of \$1. The next time we purchase cottage cheese, we remember that the first purchase seemed expensive at 50 cents an ounce. So, we decide to buy a larger carton, containing 8 ounces and costing \$2. On Figure 4.33, we can say that this purchase corresponds to the point at which our horizontal construct intercepts the $M = 8$ curve. We look down at the corresponding SNR, and call this intercept our cost of \$2. Notice that we bought a larger container so the price went up, but because we bought a greater quantity, the price per ounce went down (the unit cost is now only 25 cents per ounce). We can continue this analogy by purchasing larger and larger containers so that the price of the container (SNR) keeps going up, but the price per ounce keeps going down. This is the age-old story called the *economy of scale.* Buying larger quantities at a time is commensurate with purchasing at the wholesale level; it makes for a lower unit price. Similarly, when we use orthogonal signaling with symbols that contain more bits, we need more power (more SNR), but the requirement per bit (E_b/N_0) is reduced.

4.9 SYMBOL ERROR PERFORMANCE FOR *M*-ARY SYSTEMS ($M > 2$)

4.9.1 Probability of Symbol Error for MPSK

For large energy-to-noise ratios, the symbol error performance $P_E(M)$, for equally likely, coherently detected M-ary PSK signaling, can be expressed [7] as

$$P_E(M) \approx 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{M}\right) \tag{4.105}$$

where $P_E(M)$ is the probability of symbol error, $E_s = E_b(\log_2 M)$ is the energy per symbol, and $M = 2^k$ is the size of the symbol set. The $P_E(M)$ performance curves for coherently detected MPSK signaling are plotted versus E_b/N_0 in Figure 4.35.

The symbol error performance for differentially coherent detection of M-ary DPSK (for large E_s/N_0) is similarly expressed [7] as

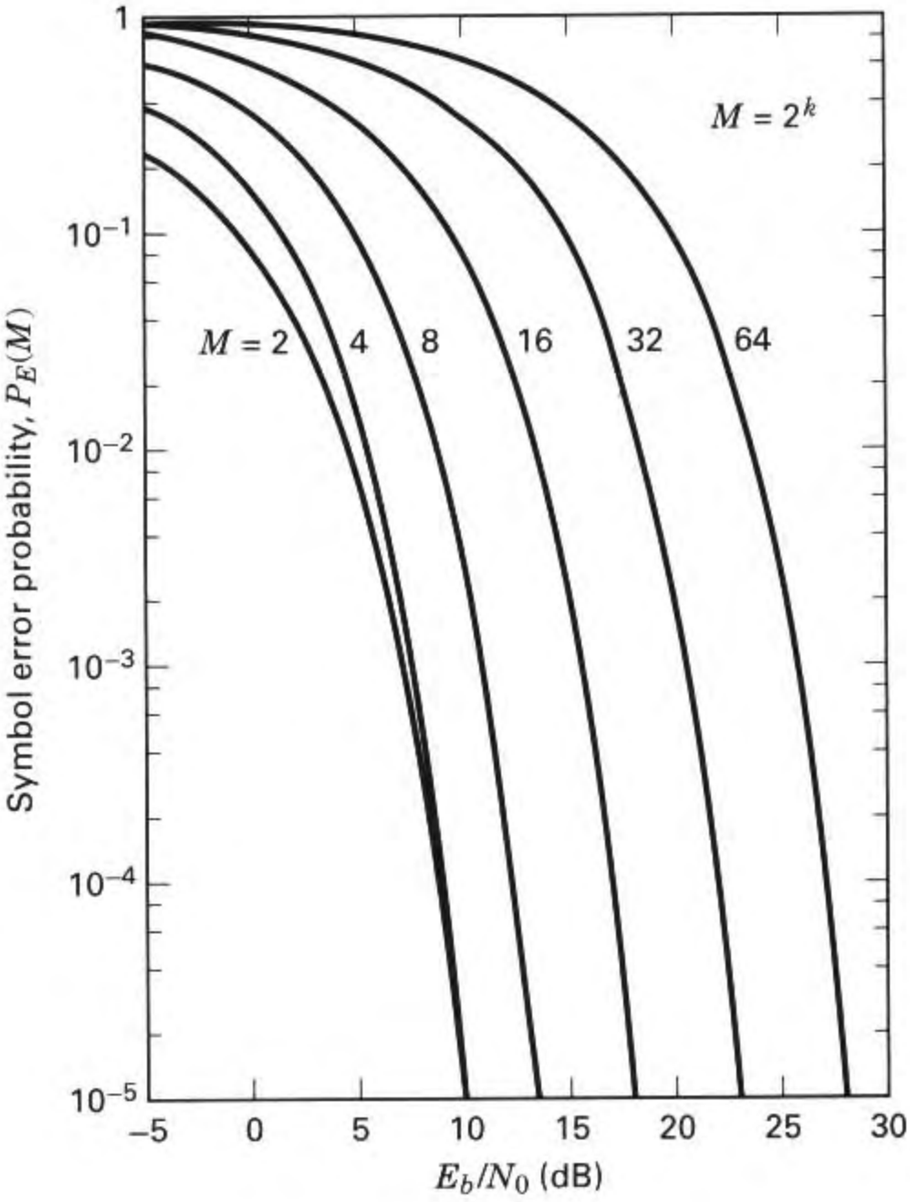


Figure 4.35 Symbol error probability for coherently detected multiple phase signaling. (Reprinted from W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering,* Prentice-Hall, Inc. Englewood Cliffs, N.J., 1973, courtesy of W. C. Lindsey and Marvin K. Simon.)

$$P_E(M) \approx 2Q\left(\sqrt{\frac{2E_s}{N_0}}\sin\frac{\pi}{\sqrt{2}M}\right) \tag{4.106}$$

4.9.2 Probability of Symbol Error for MFSK

The symbol error performance $P_E(M)$, for equally likely, *coherently* detected M-ary orthogonal signaling can be upper bounded [5] as

$$P_E(M) \leq (M-1)\,Q\left(\sqrt{\frac{E_s}{N_0}}\right) \tag{4.107}$$

where $E_s = E_b(\log_2 M)$ is the energy per symbol and M is the size of the symbol set. The $P_E(M)$ performance curves for coherently detected M-ary orthogonal signaling are plotted versus E_b/N_0 in Figure 4.36.

The symbol error performance for equally likely, *noncoherently* detected M-ary orthogonal signaling is [9]

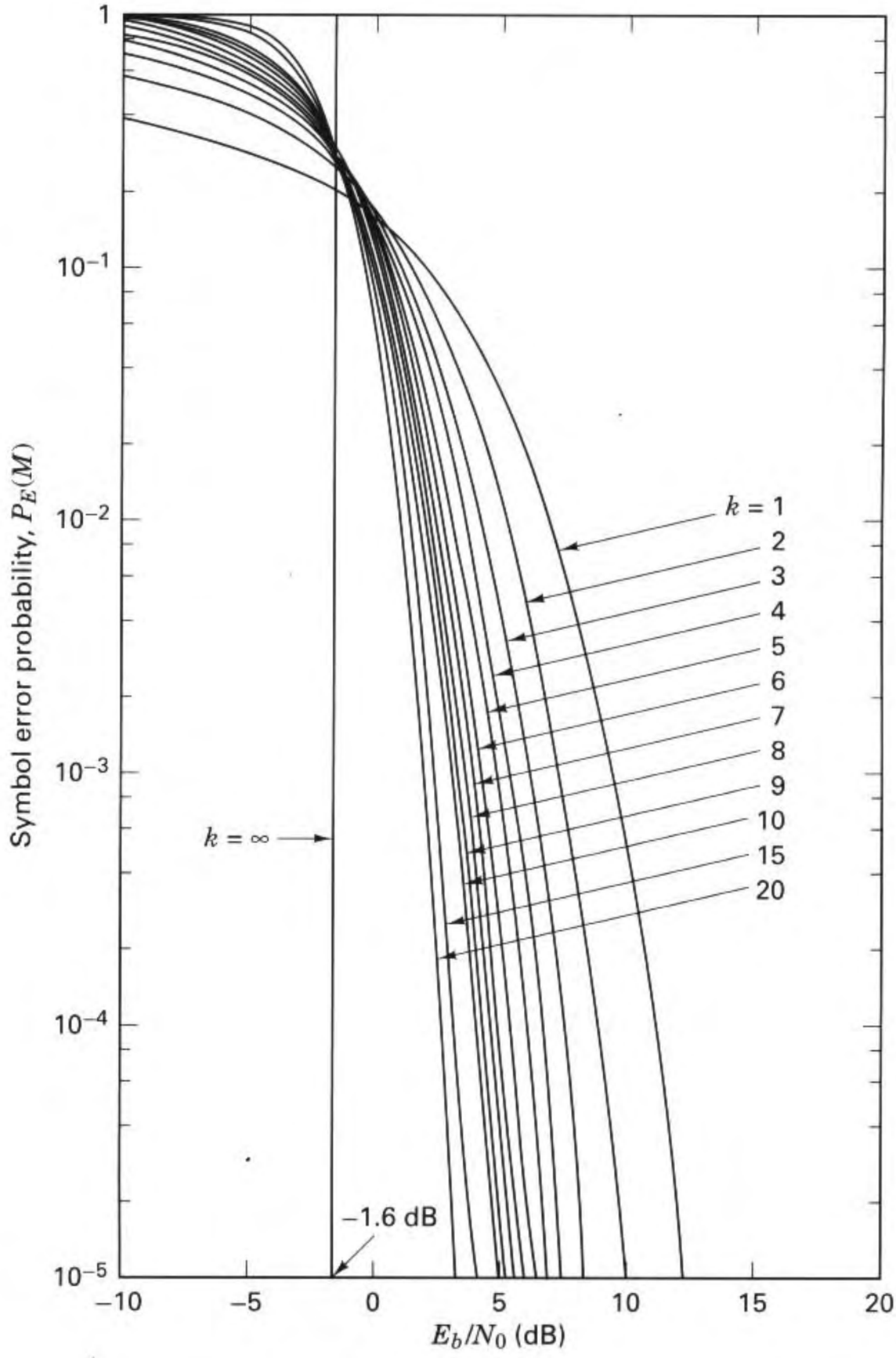


Figure 4.36 Symbol error probability for coherently detected *M*-ary orthogonal signaling. (Reprinted from W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering,* Prentice-Hall, Inc., Englewod Cliffs, N.J., 1973, courtesy of W. C. Lindsey and Marvin K. Simon.)

$$P_E(M) = \frac{1}{M}\exp\left(-\frac{E_s}{N_0}\right)\sum_{j=2}^{M}(-1)^j\binom{M}{j}\exp\left(\frac{E_s}{jN_0}\right) \tag{4.108}$$

where

$$\binom{M}{j} = \frac{M!}{j!\,(M-j)!} \tag{4.109}$$

is the standard binomial coefficient yielding the number of ways in which j symbols out of M may be in error. Note that for binary case, Equation (4.108) reduces to

$$P_B = \frac{1}{2}\exp\left(-\frac{E_b}{2N_0}\right) \tag{4.110}$$

which is the same result as that described by Equation (4.96). The $P_E(M)$ performance curves for noncoherently detected M-ary orthogonal signaling are plotted versus E_b/N_0 in Figure 4.37. If we compare this noncoherent orthogonal $P_E(M)$ performance with the corresponding $P_E(M)$ results for the coherent detection of orthogonal signals in Figure 4.36, it can be seen that for $k > 7$, there is a negligible difference. An upper bound for coherent as well as noncoherent reception of orthogonal signals is [9]

$$P_E(M) < \frac{M-1}{2}\exp\left(-\frac{E_s}{2N_0}\right) \tag{4.111}$$

where E_s is the energy per symbol and M is the size of the symbol set.

4.9.3 Bit Error Probability versus Symbol Error Probability for Orthogonal Signals

It can be shown [9] that the relationship between probability of bit error (P_B) and probability of symbol error (P_E) for an M-ary orthogonal signal set is

$$\frac{P_B}{P_E} = \frac{2^{k-1}}{2^k - 1} = \frac{M/2}{M-1} \tag{4.112}$$

In the limit as k increases, we get

$$\lim_{k\to\infty}\frac{P_B}{P_E} = \frac{1}{2}$$

A simple example will make Equation (4.112) intuitively acceptable. Figure 4.38 describes an octal message set. The message symbols (assumed equally likely) are to be transmitted on orthogonal waveforms such as FSK. With orthogonal signaling, a decision error will transform the correct signal into any one of the $(M-1)$ incorrect signals with equal probability. The example in Figure 4.38 indicates that the symbol comprising bits 0 1 1 was transmitted. An error might occur in any one of the other $2^k - 1 = 7$ symbols, with equal probability. Notice that just because a symbol error is made does not mean that all the bits within the symbol will be in error. In Figure 4.38, if the receiver decides that the transmitted symbol is the bot-

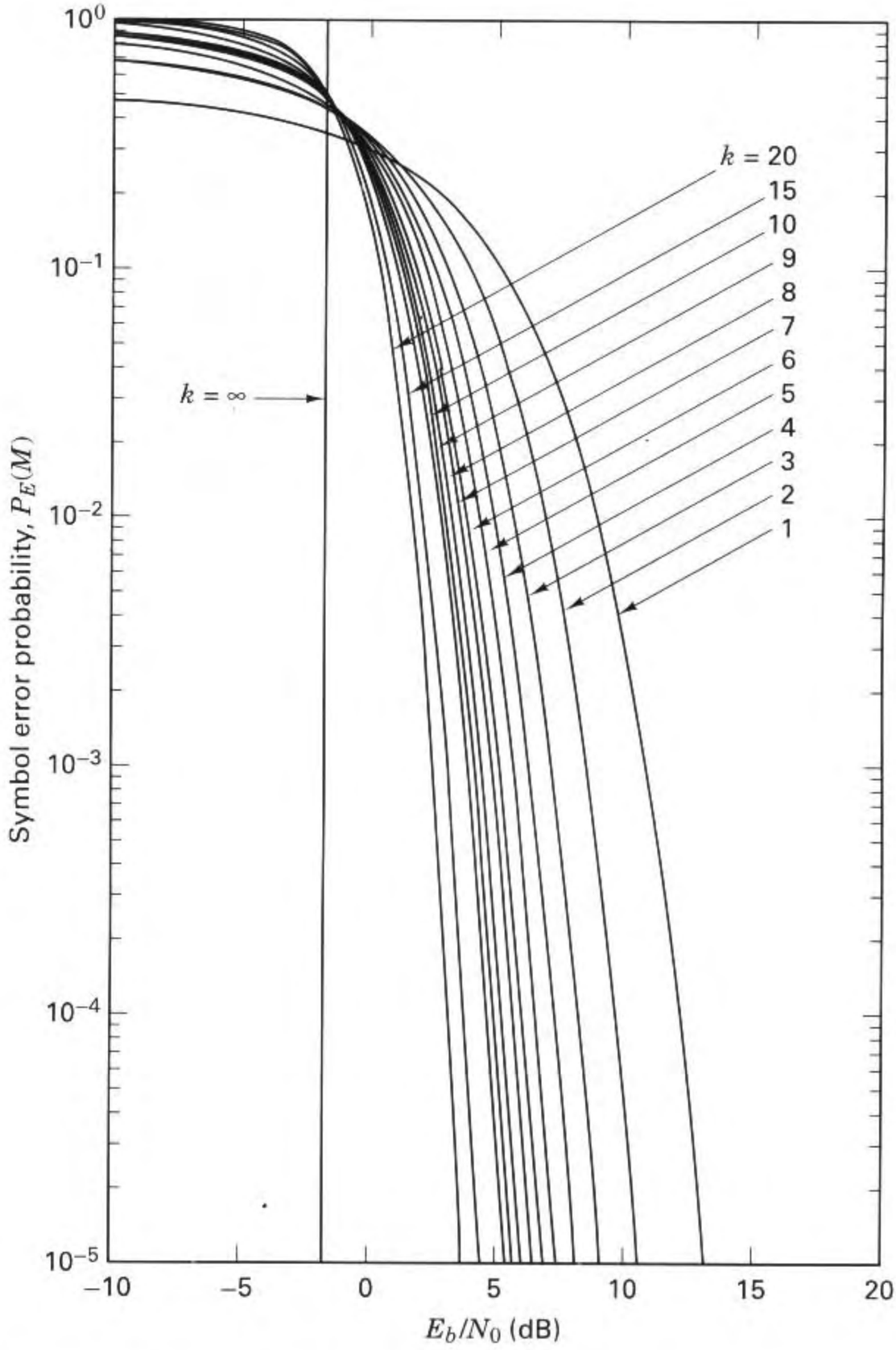


Figure 4.37 Symbol error probability for noncoherently detected *M*-ary orthogonal signaling. (Reprinted from W. C. Lindsey and M. K. Simon, *Telecommunication Systems Engineering,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973, courtesy of W. C. Lindsey and Marvin K. Simon.)

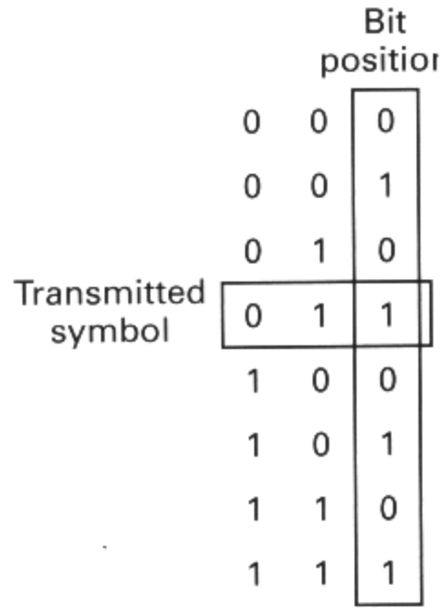


Figure 4.38 Example of P_B versus P_E.

tom one listed, comprising bits 1 1 1, two of the three transmitted symbol bits will be correct; only one bit will be in error. It should be apparent that for nonbinary signaling, P_B will always be less than P_E (keep in mind that P_B and P_E reflect the frequency of making errors on the *average*.)

Consider any of the bit-position columns in Figure 4.38. For each bit position, the digit occupancy consists of 50% ones and 50% zeros. In the context of the first bit position (rightmost column) and the transmitted symbol, how many ways are there to cause an error to the binary one? There are $2^{k-1} = 4$ ways (four places where zeros appear in the column) that a bit error can be made; it is the same for each of the columns. The final relationship P_B/P_E, for orthogonal signaling, in Equation (4.112), is obtained by forming the following ratio: the number of ways that a bit error can be made (2^{k-1}) divided by the number of ways that a symbol error can be made ($2^k - 1$). For the Figure 4.38 example, $P_B/P_E = 4/7$.

4.9.4 Bit Error Probability Versus Symbol Error Probability for Multiple Phase Signaling

For the case of MPSK signaling, P_B is less than or equal to P_E, just as in the case of MFSK signaling. However, there is an important difference. For orthogonal signaling, selecting any one of the $(M - 1)$ erroneous symbols is equally likely. In the case of MPSK signaling, each signal vector is not equidistant from all of the others. Figure 4.39a illustrates an 8-ary decision space with the pie-shaped decision regions denoted by the 8-ary symbols in binary notation. If symbol (0 1 1) is transmitted, it is clear that should an error occur, the transmitted signal will most likely be mistaken for one of its closest neighbors, (0 1 0) or (1 0 0). The likelihood that (0 1 1) would get mistaken for (1 1 1) is relatively remote. If the assignment of bits to symbols follows the binary sequence shown in the symbol decision regions of Figure 4.39a, some symbol errors will usually result in two or more bit errors, even with a large signal-to-noise ratio.

For nonorthogonal schemes, such as MPSK signaling, one often uses a binary-to-M-ary code such that binary sequences corresponding to adjacent sym-

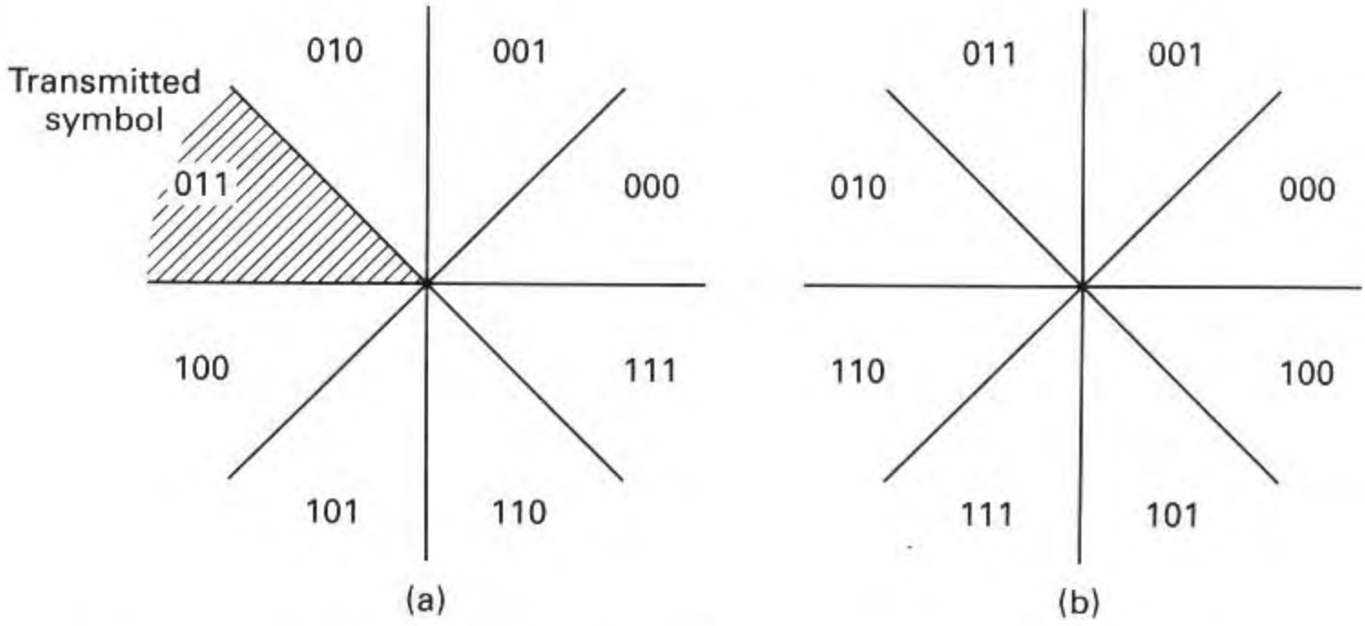


Figure 4.39 Binary-coded versus Gray-coded decision regions in an MPSK signal space. (a) Binary coded. (b) Gray coded.

bols (phase shifts) differ in only one bit position; thus when an M-ary symbol error occurs, it is more likely that only one of the k input bits will be in error. A code that provides this desirable feature is the Gray code [7]; Figure 4.39b illustrates the bit-to-symbol assignment using a Gray code for 8-ary PSK. Here it can be seen that neighboring symbols differ from one another in only one bit position. Therefore, the occurrence of a multibit error, for a given symbol error, is much reduced compared to the uncoded binary assignment seen in Figure 4.39a. Implementing such a Gray code, represents one of the few cases in digital communications where a benefit can be achieved without incurring any cost. The Gray code is simply an assignment that requires no special or additional circuitry. Utilizing the Gray code assignment, it can be shown [5] that

$$P_B \approx \frac{P_E}{\log_2 M} \qquad (\text{for } P_E \ll 1) \tag{4.113}$$

Recall from Section 4.8.4 that BPSK and QPSK signaling have the same bit error probability. Here, in Equation (4.113), we verify that they do not have the same symbol error probability. For BPSK, $P_E = P_B$. However, for QPSK, $P_E \approx 2P_B$.

An exact closed-form expression for the bit-error probability P_B of 8-ary PSK, together with tight upper and lower bounds on P_B for M-ary PSK with larger M, may be found in Lee [10].

4.9.5 Effects of Intersymbol Interference

In the previous sections and in Chapter 3 we have treated the detection of signals in the presence of AWGN under the assumption that there is no intersymbol interference (ISI). Thus the analysis has been straightforward, since the zero-mean AWGN process is characterized by its variance alone. In practice we find that ISI is

often a second source of interference which must be accounted for. As explained in Section 3.3, ISI can be generated by the use of bandlimiting filters at the transmitter output, in the channel, or at the receiver input. The result of this additional interference is to degrade the error probabilities for coherent as well as for noncoherent reception. Calculating error performance in the presence of ISI in addition to AWGN is much more complicated since it involves the impulse response of the channel. The subject will not be treated here; however, for those readers interested in the details of the analysis, References [11–16] should prove interesting.

4.10 CONCLUSION

We have catalogued some basic bandpass digital modulation formats, particularly phase shift keying (PSK) and frequency shift keying (FSK). We have considered a geometric view of signal vectors and noise vectors, particularly antipodal and orthogonal signal sets. This geometric view allows us to consider the detection problem in the light of an orthogonal signal space and signal regions. This view of the space, and the effect of noise vectors causing transmitted signals to be received in the incorrect region, facilitates the understanding of the detection problem and the performance of various modulation and demodulation techniques. In Chapter 9 we reconsider the subjects of modulation and demodulation, and we investigate some bandwidth-efficient modulation techniques.

REFERENCES

1. Schwartz, M., *Information, Transmission, Modulation, and Noise,* McGraw-Hill Book Company, New York, 1970.
2. Van Trees, H. L., *Detection, Estimation, and Modulation Theory,* Part 1, John Wiley & Sons, Inc., New York, 1968.
3. Park, J. H., Jr., "On Binary DPSK Detection," *IEEE Trans. Commun.,* vol. COM26, no. 4, Apr. 1978, pp. 484–486.
4. Ziemer, R. E., and Peterson, R. L., *Digital Communications and Spread Spectrum Systems,* Macmillan Publishing Company, Inc., New York, 1985.
5. Lindsey, W. C., and Simon, M. K., *Telecommunication Systems Engineering,* Prentice-Hall, Inc., Englewood Cliffs, N.J., 1973.
6. Whalen, A. D., *Detection of Signals in Noise,* Academic Press, Inc., New York, 1971.
7. Korn, I., *Digital Communications,* Van Nostrand Reinhold Company, Inc., New York, 1985.
8. Couch, L. W. II, *Digital and Analog Communication Systems,* Macmillan Publishing Company, New York, 1983.
9. Viterbi, A. J., *Principles of Coherent Communications,* McGraw-Hill Book Company, New York, 1966.
10. Lee, P. J., "Computation of the Bit Error Rate of Coherent M-*ary* PSK with Gray Code Bit Mapping," *IEEE Trans. Commun.,* vol. COM34, no. 5, May 1986, pp. 488–491.

11. Hoo, E. Y., and Yeh, Y. S., "A New Approach for Evaluating the Error Probability in the Presence of the Intersymbol Interference and Additive Gaussian Noise," *Bell Syst. Tech. J.*, vol. 49, Nov. 1970, pp. 2249–2266.
12. Shimbo, O., Fang, R. J., and Celebiler, M., "Performance of *M*-ary PSK Systems in Gaussian Noise and Intersymbol Interference," *IEEE Trans. Inf. Theory,* vol. IT19, Jan. 1973, pp. 44–58.
13. Prabhu, V. K., "Error Probability Performance of *M*-ary CPSK Systems with Intersymbol Interference," *IEEE Trans. Commun.,* vol. COM21, Feb. 1973, pp. 97–109.
14. Yao, K., and Tobin, R. M., "Moment Space Upper and Lower Error Bounds for Digital Systems with Intersymbol Interference," *IEEE Trans. Inf. Theory,* vol. IT22, Jan. 1976, pp. 65–74.
15. King, M. A., Jr., "Three Dimensional Geometric Moment Bounding Techniques," *J. Franklin Inst.,* vol. 309, no. 4, Apr. 1980, pp. 195–213.
16. Prabhu, V. K., and Salz, J., "On the Performance of Phase-Shift Keying Systems," *Bell Syst. Tech. J.,* vol. 60, Dec. 1981, pp. 2307–2343.

PROBLEMS

4.1. Find the expected number of bit errors made in one day by the following continuously operating coherent BPSK receiver. The data rate is 5000 bits/s. The input digital waveforms are $s_1(t) = A \cos \omega_0 t$ and $s_2(t) = -A \cos \omega_0 t$ where $A = 1$ mV and the single-sided noise power spectral density is $N_0 = 10^{-11}$ W/Hz. Assume that signal power and energy per bit are normalized relative to a 1-Ω resistive load.

4.2. A continuously operating coherent BPSK system makes errors at the average rate of 100 errors per day. The data rate is 1000 bits/s. The single-sided noise power spectral density is $N_0 = 10^{-10}$ W/Hz.

(a) If the system is ergodic, what is the average bit error probability?

(b) If the value of received average signal power is adjusted to be 10^{-6} W, will this received power be adequate to maintain the error probability found in part (a)?

4.3. If a system's main performance criterion is bit error probability, which of the following two modulation schemes would be selected for an AWGN channel? Show computations.

Binary noncoherent orthogonal FSK with $E_b/N_0 = 13$ dB

Binary coherent PSK with $E_b/N_0 = 8$ dB

4.4. The bit stream

1 0 1 0 1 0 1 1 1 1 0 1 0 1 0 1 0 0 0 0 1 1 1 1

is to be transmitted using DPSK modulation. Show four different differentially encoded sequences that can represent the data sequence above, and explain the algorithm that generated each.

4.5. (a) Calculate the minimum required bandwidth for a noncoherently detected orthogonal binary FSK system. The higher-frequency signaling tone is 1 MHz and the symbol duration is 1 ms.

(b) What is the minimum required bandwidth for a noncoherent MFSK system having the same symbol duration?

4.6. Consider a BPSK system with equally likely waveforms $s_1(t) = \cos \omega_0 t$ and $s_2(t) = -\cos \omega_0 t$. Assume that the received $E_b/N_0 = 9.6$ dB, giving rise to a bit-error probability of 10^{-5}, when the synchronization is perfect. Consider that carrier recovery with the PLL suffers some fixed error ϕ associated with the phase estimate, so that the reference signals are expressed as $\cos(\omega_0 t + \phi)$ and $-\cos(\omega_0 t + \phi)$. Note that the error-degradation effect of a fixed known bias can be computed by using the closed-end relationships presented in this chapter. However, if the phase error were to consist of a random jitter, computing its effect would then require a stochastic treatment. (See Chapter 10.)

(a) How badly does the bit-error probability degrade when $\phi = 25°$?

(b) How large a phase error would cause the bit-error probability to degrade to 10^{-3}?

4.7. Find the probability of bit error, P_B, for the coherent matched filter detection of the equally likely binary FSK signals

$$s_1(t) = 0.5 \cos 2000\pi$$

and

$$s_2(t) = 0.5 \cos 2020\pi t,$$

where the two-sided AWGN power spectral density is $N_0/2 = 0.0001$. Assume that the symbol duration is $T = 0.01$ s.

4.8. Find the optimum (minimum probability of error) threshold γ_0, for detecting the equally likely signals $s_1(t) = \sqrt{2E/T} \cos \omega_0 t$ and $s_2(t) = \sqrt{\frac{1}{2}E/T} \cos(\omega_0 t + \pi)$ in AWGN, using a correlator receiver as shown in Figure 4.7b. Assume a reference signal of $\psi_1(t) = \sqrt{2/T} \cos \omega_0 t$.

4.9. A system using matched filter detection of equally likely BPSK signals, $s_1(t) = \sqrt{2E/T} \cos \omega_0 t$ and $s_2(t) = \sqrt{2E/T} \cos(\omega_0 t + \pi)$, operates in AWGN with a received E_b/N_0 of 6.8 dB. Assume that $\mathbf{E}\{z(T)\} = \pm \sqrt{E}$.

(a) Find the minimum probability of bit error, P_B, for this signal set and E_b/N_0.

(b) If the decision threshold is $\gamma = 0.1 \sqrt{E}$, find P_B.

(c) The threshold of $\gamma = 0.1 \sqrt{E}$ is optimum for a particular set of a priori probabilities, $P(s_1)$ and $P(s_2)$. Find the values of these probabilities (refer to Section B.2).

4.10. (a) Describe the impulse response of a matched filter for detecting the discrete signal shown in Figure P4.1. With this signal at the input to the filter, show the output as a function of time. Neglect the effects of noise. What is the maximum output value?

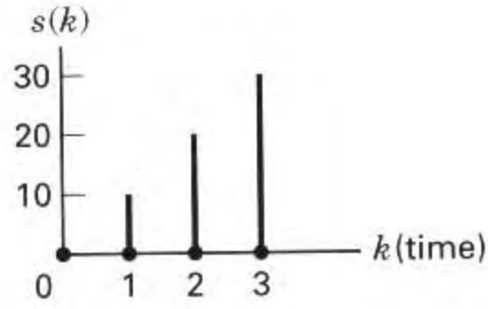


Figure P4.1

(b) In a matched filter (MF), a signal is convolved with a time-reversed function of the signal (impulse response of the MF). Convolution reverses the function again; thus, the MF yields the correlation of a signal and its "look alike" copy (even though the MF operation is designated as convolution). In implementing a

MF, suppose that you accidently connect the circuitry so as to yield the correlation of a signal and its time-reversed copy; the output would then yield a signal convolved with itself. Show the output as a function of time. What is the maximum output value? Note that the maximum output value for part (a) will occur at a different time index compared with that of part (b).

(c) By examining the filter's output vales from the flawed circuit of part (b), compared with the correct values in part (a), can you find a clue that can help you predict when such an output sequence appears to be a valid matched filter output, and when it does not?

(d) If noise were added to the signal, compare the output SNR for the correlator versus the convolver. Also, if the input consists of noise only, compare the output of the correlator versus the convolver.

4.11. A binary source with equally likely symbols controls the switch position in a transmitter operating over an AWGN channel, as shown in Figure P4.2. The noise has two-sided spectral density $N_0/2$. Assume antipodal signals of time duration T seconds and energy E joules. The system clock produces a clock pulse every T seconds, and the binary source rate is $1/T$ bits/s. Under *normal* operation, the switch is up when the source produces a binary zero, and it is down when the source produces a binary one. However, the switch is *faulty*. With probability p, it will be thrown in the wrong direction during a given T-second interval. The presence of a switch error during any interval is independent of the presence of a switch error at any other time. Assume that $\mathbf{E}\{z(T)\} = \pm\sqrt{E}$.

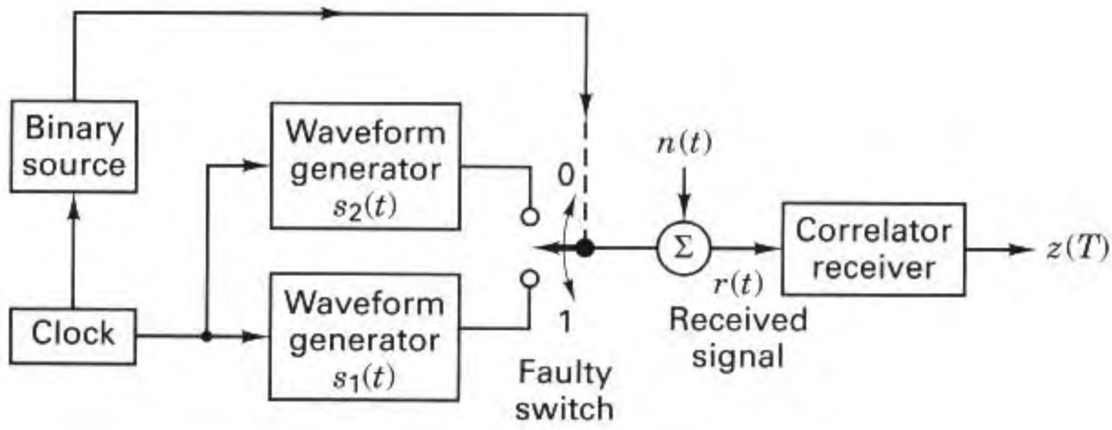


Figure P4.2

(a) Sketch the conditional probability functions, $p(z|s_1)$ and $p(z|s_2)$.

(b) The correlator receiver observes $r(t)$ in the interval $(0, T)$. Sketch the block diagram of an optimum receiver for minimizing the bit error probability when it is known that the switch is faulty with probability, p.

(c) Which one of the following two systems would you prefer to have?

$$p = 0.1 \text{ and } \frac{E_b}{N_0} = \infty$$

or

$$p = 0 \text{ and } \frac{E_b}{N_0} = 7 \text{ dB}$$

4.12. **(a)** Consider a 16-ary PSK system with symbol error probability $P_E = 10^{-5}$. A Gray code is used for the symbol to bit assignment. What is the approximate bit error probability?

(b) Repeat part (a) for a 16-ary orthogonal FSK system.

4.13. Consider a coherent orthogonal MFSK system with $M = 8$ having the equally likely waveforms $s_i(t) = A \cos 2\pi f_i t$, $i = 1, \ldots, M$, $0 \le t \le T$, where $T = 0.2$ ms. The received carrier amplitude, A, is 1 mV, and the two-sided AWGN spectral density, $N_0/2$, is 10^{-11} W/Hz. Calculate the probability of bit error, P_B.

4.14. A bit error probability of $P_B = 10^{-3}$ is required for a system with a data rate of 100 kbits/s to be transmitted over an AWGN channel using coherently detected MPSK modulation. The system bandwidth is 50 kHz. Assume that the system frequency transfer function is a raised cosine with a roll-off characteristic of $r = 1$ and that a Gray code is used for the symbol to bit assignment.

(a) What E_s/N_0 is required for the specified P_B?

(b) What E_b/N_0 is required?

4.15. A differentially coherent MPSK system operates over an AWGN channel with an E_b/N_0 of 10 dB. What is the symbol error probability for $M = 8$ and equally likely symbols?

4.16. If a system's main performance criterion is bit-error probability, which of the following two modulation schemes would be selected for transmission over an AWGN channel?

$$\text{coherent 8-ary orthogonal FSK with } \frac{E_b}{N_0} = 8 \text{ dB}$$

or

$$\text{coherent 8-ary PSK with } \frac{E_b}{N_0} = 13 \text{ dB}$$

(Assume that a Gray code is used for the MPSK symbol-to-bit assignment, and show computations.)

4.17. Consider that a BPSK demodulator/detector has a synchronization error consisting of a time bias pT, where p is a fraction ($0 \le p \le 1$) of the symbol time T. In other words, the detection of a symbol starts early (late) and concludes early (late) by an amount pT. Assume equally likely signaling and perfect frequency and phase synchronization. Note that the error-degradation effect of a fixed known bias can be computed by using the closed-end relationships presented in this chapter. However, if the timing error were to consist of a random jitter, computing its effect would then require a stochastic treatment. (See Chapter 10.)

(a) Find the general expression for bit-error probability P_b as a function of p.

(b) If the received $E_b/N_0 = 9.6$ dB and $p = 0.2$, compute the value of degraded P_b due to the timing bias.

(c) If one did not compensate for the timing bias in this example, how much additional E_b/N_0 in dB must be provided in order to restore the P_b that exists when $p = 0$?

4.18. Using all of the stated specifications, repeat problem 4.17 for the case of coherently detected, binary frequency-shift-keying (BFSK) modulation.

4.19. Consider that a BPSK demodulator/detector has a synchronization error consisting of a time bias pT, where p is a fraction ($0 \le p \le 1$) of the symbol time T. Consider that

there is also a constant phase-estimation error ϕ. Assume equally-likely signaling and perfect frequency synchronization.

(a) Find the general expression for bit-error probability P_b as a function of p and ϕ.

(b) If received $E_b/N_0 = 9.6$ dB, $p = 0.2$, and $\phi = 25°$, compute the value of degraded P_b due to the combined effects of timing and phase bias.

(c) If one did not compensate for the biases in this example, how much additional E_b/N_0 in dB must be provided in order to restore the P_b that exists when $p = 0$ and $\phi = 0°$?

4.20. Correlating to a known Barker sequence is an often used synchronization technique, since the Barker sequence yields a prominent correlation peak when properly synchronized, and a small correlation output when not synchronized. Using the short Barker sequence 1 0 1 1 1, where the leftmost bit is the earliest bit, devise a discrete matched filter similar to the one in Figure 4.10 that is matched to this sequence. Verify its usefulness for synchronization by plotting the output versus input as a function of time, when the input is the 1 0 1 1 1 sequence.

QUESTIONS

4.1. At what location in the system is E_b/N_0 defined? (See Section 4.3.2.)

4.2. Amplitude- or phase-shift keying is visualized as a constellation of points or phasors on a plane. Why can't we use a similarly simple visualization for orthogonal signaling such as FSK? (See Section 4.4.4.)

4.3. In the case of MFSK signaling, what is the minimum tone spacing that insures signal *orthogonality?* (See Section 4.5.4.)

4.4. What benefits are there in using *complex notation* for representing sinusoids? (See Sections 4.2.1 and 4.6.)

4.5. Digital modulation schemes fall into one of the two classes with opposite behavior characteristics: *orthogonal* signaling, and *phase/amplitude* signaling. Describe the behavior of each class. (See Sections 4.8.2.)

4.6. Why do binary phase shift keying (BPSK) and quaternary phase shift keying (QPSK) manifest the same bit-error-probability relationship? (See Section 4.8.4.)

4.7. In the case of multiple-phase shift keying (MPSK), why does *bandwidth efficiency* improve with higher dimensional signaling? (See Sections 4.8.2 and 4.8.3.)

4.8. In the case of orthogonal signaling such as MFSK, why does *error-performance* improve with higher dimensional signaling? (See Section 4.8.5.)

4.9. The use of a Gray code for assigning bits to symbols, represents one of the few cases in digital communications where a benefit can be achieved *free-of-charge.* Explain why there is no cost. (See Section 4.9.4.)

EXERCISES

Using the Companion CD, run the exercises associated with Chapter 4.